Microeconomic Theory I: Math

Terence Johnson
tjohns20@nd.edu
nd.edu/∼tjohns20/micro601.html

Updated 11/26/2012

Contents

Chapter 1

Introduction

If you learn everything in this math handout — text and exercises — you will be able to work through most of MWG and pass comprehensive exams with reasonably high likelihood. The math presented here is, honestly, what I think you really, truly need to know. I think you need to know these things not just to pass comps, but so that you can competently skim an important Econometrica paper to learn a new estimator you need for your research, or to pick up a book on solving non-linear equations numerically written by a non-economist and still be able to digest the main ideas. Tools like Taylor series come up in asymptotic analysis of estimators, proving necessity and sufficiency of conditions for a point to maximize a function, and the theory of functional approximation, which touch every field from applied micro to empirical macro to theoretical micro. Economists use these tools are used all the time, and not learning them will handicap you in your ability to continue acquiring skills.

1.1 Economic Models

There are five ingredients to an economic model:

- Who are the agents? (Players)
- What can the agents decide to do, and what outcomes arise as a consequence of the agents' choices? (Actions)
- What are the agents' preferences over outcomes? (Payoffs)
- When do the agents make decisions? (Timing)
- What do the agents know when they decide how to act? (Information)

Once the above ingredients have been fixed, the question then arises how agents will behave. Economics has taken the hard stance that agents will act deliberately to influence the outcome in their favor, or that each agent maximizes his own payoff. This does not mean that agents cannot care — either for selfish reasons, or per se — about the welfare or happiness of other agents in the economy. It simply means that, when faced with a decision, agents act deliberately to get the best outcome available to them according to their own preferences, and not as, say, a wild animal in the grip of terror (such a claim is controversial, at least in the social sciences). Once the model has been fixed, the rest of the analysis is largely an application of mathematical and statistical reasoning.

1.2 Perfectly Competitive Markets

The foundational economic model is the classical "Marshallian" or "partial equilibrium" market. This is the model you are probably most familiar with from undergraduate economics. In a price-taking market, there are two agents: a representative firm and a representative consumer.

The firm's goal is to maximize profits. It chooses how much of its product to make, q, taking as given the price, p, and its costs, $C(q)$. Then the firm's profits are

$$\pi(q) = pq - C(q)$$

The consumer's goal is to maximize his utility. The consumer chooses how much quantity to purchase, q, taking as given the price, p, and wealth, w. The consumer's utility function takes the form

$$u(q, m)$$

where m is the amount of money spent on goods other than q. The consumer's utility function is *quasi-linear* if

$$u(q, m) = v(q) + m$$

where $v(q)$ is positive, increasing, and has diminishing marginal benefit. Then the consumer is trying to solve

$$\max_{q,m} v(q) + m$$

subject to a budget constraint $w = pq + m$.

An outcome is an *allocation* (q^*, m^*) of goods giving the amount traded between the consumer and the firm, q^* and the amount of money the consumer spends on other goods, m^*.

The allocation will be decided by finding the *market-clearing price and quantity*. The consumer is asked to report how much quantity they demand $q^D(p)$ at each price p, and the firm is asked how much quantity $q^S(p)$ it is willing to supply at each price p. The market clears where $q^D(p^*) = q^S(p^*) = q^*$, which is the market-clearing quantity; the market-clearing price is p^*.

The market meets once, and all of the information above is known by all the agents.

Then the model is

- Agents: The consumer and firm.
- Actions: The consumer reports a demand curve $q^D(p)$, and the firm report a supply curve $q^S(p)$, both taking the price as given.
- Payoffs: The consumer and firm trade the market-clearing quantity at the market-clearing price, giving the consumer a payoff of $u(q^*, w - p^*q^*)$ and the firm a payoff of $p^*q^* - C(q^*)$.
- Information: The consumer and firm both know everything about the market.
- Timing: The market meets once, and the demand and supply curves are submitted simultaneously.

1.2.1 Price-Taking Equilibrium

Let's make further assumptions so that the model is extremely easy to solve. First, let $C(q) = \dfrac{c}{2}q^2$, and let $v(q) = b\log(q)$.

The Supply Side

Since firms are maximizing profits, their objective is

$$\max_q pq - \frac{c}{2}q^2$$

The first-order necessary condition for the firm is

$$p - cq^S = 0$$

and the second-order sufficient condition for the firm is

$$-c < 0$$

which is automatically satisfied. Solving the FONC gives the *supply curve*,

$$q^S(p) = \frac{p}{c}$$

This expresses the firm's willingness to produce the good as a function of their costs, parametrized by c, and the price they receive, p.

The Demand Side

Consumers act to maximize utility subject to their budget constraint, or

$$\max_q b\log(q) + m$$

subject to $w = pq + m$. To handle the constraint, re-write it as $m = w - pq$ and substitute it into the objective function to get

$$\max_q b\log(q) + w - pq$$

Then the first-order necessary condition for the consumer is

$$\frac{b}{q^D} - p = 0$$

and the second-order sufficient condition for the consumer is

$$-\frac{1}{(q^D)^2} < 0$$

which is automatically satisfied. Solving the FONC gives the *demand curve*,

$$q^D(p) = \frac{b}{p}$$

This expresses how much quantity the consumer would like to purchase as a function of preferences b and the price p.

Market-Clearing Equilibrium

In equilibrium, supply equals demand, and the market-clearing price p^* and market-clearing quantity q^* are determined as

$$q^D(p^*) = q^S(p^*) = q^*$$

or

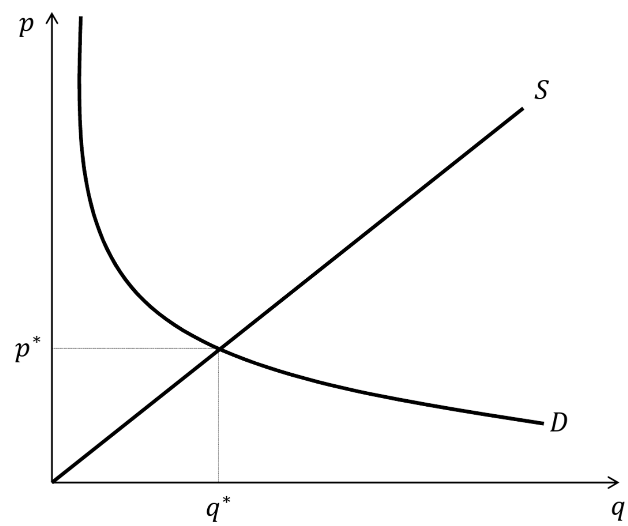
$$\frac{b}{p^*} = \frac{p^*}{c} = q^*$$

Solve for p^* yields

$$p^* = \sqrt{bc}$$

and

$$q^* = \frac{\sqrt{b}}{\sqrt{c}}$$



Market Equilibrium

Notice that the variables of interest — p^* and q^* — are expressed completely in terms of the parameters of the model that are outside of any agents' control, b and c. We can now vary b and c freely to see how the equilibrium price and quantity vary, and study how changes in tastes or technology will change behavior.

1.3 Basic questions of economic analysis

After building a model, two *mathematical* questions naturally arise,

- How do we know solutions exist to the agents' maximization problems? (the Weierstass theorem)
- How do we find solutions? (first order necessary conditions, second order sufficient conditions)

as well as two *economic* questions,

- How does an agent's behavior respond to changes in the economic environment? (the implicit function theorem)

- How does an agent's payoff respond to changes in the economic environment? (the envelope theorem)

In Micro I, we learn in detail the nuts and bolts of solving agents' problems in isolation (optimization theory). In Micro II, you learn — through general equilibrium and game theory — how to solve for the behavior when many agents are interacting at once (equilibrium and fixed-point theory).

The basic methodology for solving problems in Micro I is:

- Check that a solution exists to the problem using the Weierstrass theorem
- Build a candidate list of potential solutions using the appropriate first-order necessary conditions
- Find the global maximizer by comparing candidate solutions directly by computing their payoff, or use the appropriate second-order sufficient conditions to verify which are maxima
- Study how an agent's behavior changes when economic circumstances change using the implicit function theorem
- Study how an agent's payoffs changes when economic circumstances change using the envelope theorem

Actually, the above is a little misleading. There are many versions of the Weierstrass theorem, the FONCs, the SOSCs, the IFT and the envelope theorem. The appropriate version depends on the kind of problem you face: Is there a single choice variable or many? Are there equality or inequality constraints? And so on.

The entire class — including the parts of Mas-Colell-Whiston-Green that we cover — is nothing more than the application of the above algorithm to specific problems. At the end of the course, it is imperative that you understand the above algorithm and know how to implement it to pass the comprehensive exams.

We will start by studying the algorithm in detail for one-dimensional maximization problems in which agents only choose one variable. Then we will generalize it to multi-dimensional maximization problems in which agents choose many things at once.

Exercises

The exercises refer to the simple partial equilibrium model of Section 1.2.

1. [Basics] As b and c change, how do the supply and demand curves shift? How do the equilibrium price and quantity change? Sketch graphs as well as compute derivatives.

2. [Taxes] Suppose there is a tax t on the good q. A portion μ of t is paid by the consumer for each unit purchased, and a portion $1-\mu$ of t is paid by the firm for each unit sold. How does the consumer's demand function depend on μ? How does the firm's supply function depend on μ? How does the equilibrium price and quantity p^* and q^* depend on μ? If t goes up, how are the market clearing price and quantity affected? Sketch a graph of the equilibrium in this market.

3. [Long-run equilibrium] (i) Suppose that there are K firms in the industry with short-run cost function $C(q) = \frac{c}{2}q^2$, so that each firm produces q, but *aggregate supply* is $Kq = Q$. Consumers maximize $b\log(Q) + m$ subject to $w = pQ + m$. Solve for the market-clearing price and quantity for each K and the short-run profits of the firms. (ii) Suppose firms have to pay a fixed cost F to enter the industry. Find the equilibrium number of firms in the long run, K^*, if there is entry as

long as profits are strictly positive. How does K^* vary in F, b and c? How do the market-clearing price and quantity vary in the long run with F?

4. [Firm cost structure] Suppose a price-taking firm's costs are $C(q) = c(q) + F$, where $c(0) = 0$, $c'(0) > 0$ and $c''(0) > 0$ and F is a fixed cost. (i) Show that marginal cost $C'(q)$ intersects average cost $C(q)/q$ at the minimum of $C(q)/q$. This is the efficient scale. (ii) How does the firm's optimal choice of q depend on F?

5. [Monopoly] Suppose now that the consumers' utility function is

$$2b\sqrt{q} + m$$

and the budget constraint is $w = pq + m$. Suppose there is a single firm with cost function $C(q) = cq$. (i) Derive the demand curve $q^D(p)$ and inverse demand curve $p^D(q)$. (ii) If the monopolist recognizes its influence on the market-clearing price and quantity, it will maximize

$$\max_q p^D(q)q - cq$$

or

$$\max_p q^D(p)(p - c)$$

Show that the solutions to these problems are the same. If total revenue is $pq^D(p)$ show that its derivative, the marginal revenue curve, lies below the total revenue curve, and compare the monopolist's FONC with a price-taking firm's FONC in the same market.

6. [Efficiency] A benevolent, utilitarian social planner (or well-meaning government) would choose the market-clearing quantity to maximize the sum of the two agents' payoffs, or

$$\max_{p,q}(v(q) + w - pq) + (pq - C(q))$$

Show that this outcome is the same as that selected by the perfectly competitive market. Conclude that a competitive equilibrium achieves the same outcome that a benevolent government would pick. Give an argument for why a government trying to intervene in a decentralized market would then probably achieve a worse outcome. Give an argument why a decentralized market would probably achieve a worse outcome than a well-meaning government. Show that the allocation selected by the decentralized market and utilitarian social planner is *not* the allocation selected by a monopolist. Sketch a graph of the situation.

Congratulations, you are now qualified to be a micro TA!

Chapter 2

Basic Math and Logic

These are basic definitions that appear in all of mathematics and modern economics, but reviewing it doesn't hurt.

2.1 Set Theory

Definition 2.1.1 *A* set *A is a collection of* elements*. If x is an* element *or* member *of A, then $x \in A$. If x is not in A, then $x \notin A$. If all the elements of A and B are the same, $A = B$. The set with no elements is called* the empty set*, $\varnothing$.*

We often enumerate sets by collecting the elements between braces, as

$$F = \{ \text{ Apples, Bananas, Pears } \}$$

$$F' = \{ \text{ Cowboys, Bears, Chargers, Colts} \}$$

We can build up and break down sets as follows:

Definition 2.1.2 *Let A, B and C be sets.*

- *A is a* subset *of B if all the elements of A are elements of B, written $A \subseteq B$; if there exists an element $x \in A$ but $x \notin B$, we can also write $A \subset B$.*
- *If $A \subseteq B$ and $B \subseteq A$, then $A = B$.*
- *The set of all elements in either A or B is the* union *of A and B, written $A \cup B$.*
- *The set of all elements in both A and B is the* intersection *of A and B, written $A \cap B$. If $A \cap B = \varnothing$, then A and B are* disjoint*.*
- *Suppose $A \subset B$. The set of elements in B but not in A is the* complement of A in B*, written A^c.*

These precisely define all the normal set operations like union and intersection, and give exact meaning to the symbol $A = B$. It's easy to see that operations like union and intersection are associative (check) and commutative (check). But how does taking the complement of a union or intersection behave?

Theorem 2.1.3 (DeMorgan's Laws) *For any sets A and B that are both subsets of X,*

$$(A \cup B)^c = A^c \cap B^c$$

and

$$(A \cap B)^c = A^c \cup B^c,$$

where the complement is taken with respect to X.

Proof Suppose $x \in (A \cup B)^c$. Then x is not in the union of A and B, so x is in neither A nor B. Therefore, x must be in the complement of both A and B, so $x \in A^c \cap B^c$. This shows that $(A \cup B)^c \subset A^c \cap B^c$.

Suppose $x \in A^c \cap B^c$. Then x is contained in the complement of both A and B, so it not a member of either set, so it is not a member of the union $A \cup B$; that implies $x \in (A \cup B)^c$. This shows that $(A \cup B)^c \supset A^c \cap B^c$. ∎

If you ever find yourself confused by a complicated set relation $((A \cup B \cap C)^c \cup D...)$, draw a Venn diagram and try writing in words what is going on.

We are interested in defining relationships between sets. The easiest example is a function, $f : X \to Y$, which assigns a unique element of Y to at least some elements of X. However, in economics we will have to study some more exotic objects, so it pays to start slowly in defining functions, correspondences, and relations.

Definition 2.1.4 *The* (Cartesian) product *of two non-empty sets X and Y is the set of all ordered pairs $(x, y) = \{x \in X, y \in Y\}$.*

The product of two sets X and Y is often written $X \times Y$. If it's just one set copied over and over, $X \times X \times ... \times X = X^n$, and an ordered pair would be $(x_1, x_2, ..., x_n)$. For example $\mathbb{R} \times \mathbb{R} = \mathbb{R}^2$ is product of the real numbers with itself, which is just the plane. If we have a "set of sets", $\{X_i\}_{i=1}^N$, we often write $\times_{i=1}^N X_i$ or $\prod_{i=1}^N X_i$.

Example Suppose two agents, A and B, are playing the following game: If A and B both choose "heads", or HH, or both choose "tails", or TT, agent B pays agent A a dollar. If one player chooses tails and the other chooses heads, and vice versa, agent A pays agent B a dollar. Then A and B both have the strategy sets $S_A = S_B = \{H, T\}$, and an outcome is an ordered pair $(s_A, s_B) \in S_A \times S_B$. This game is called *matching pennies.*

Besides a Cartesian product, we can also take a space and investigate all of its subsets:

Definition 2.1.5 *The* power set *of A is the set of all subsets of A, often written 2^A, since it has $2^{|A|}$ members.*

For example, take $A = \{a, b\}$. Then the power set of A is $\varnothing$, $\{a, b\}$, $\{a\}$, and $\{b\}$.

2.2 Functions and Correspondences

Definition 2.2.1 *Let X and Y be sets. A* function *is a rule that assigns a single element $y \in Y$ to each element $x \in X$; this is written $y = f(x)$ or $f : X \to Y$. The set X is called the* domain*, and the set Y is called the* range*. Let $A \subset X$; then the* image of A *is the set $f(A) = \{y : x \in A, y = f(x)\}$.*

Definition 2.2.2 *A function $f : X \to Y$ is* injective *or* one-to-one *if for every $x_1, x_2 \in X$, $x_1 \neq x_2$ implies $f(x_1) \neq f(x_2)$. A function $f : X \to Y$ is* surjective *or* onto *if for all $y \in Y$, there exists an $x \in X$ such that $f(x) = y$.*

Definition 2.2.3 *A function is* invertible *if for every $y \in Y$, there is a unique $x \in X$ for which $y = f(x)$.*

Note that the definition of invertible is extremely precise: For *every* $y \in Y$, there is a *unique* $x \in X$ for which $y = f(x)$.

Do not confuse whether or not a function is invertible with the following idea:

Definition 2.2.4 *Let $f : X \to Y$, and $I \subset Y$. Then the* inverse image of I under f *is the set of all $x \in X$ such that $f(x) = y$ and $y \in I$.*

Example On $[0, \infty)$, the function $f(x) = x^2$ is invertible, since $x = \sqrt{y}$ is the unique inverse element. Then for any set $I \subset [0, \infty)$, the inverse image $f^{-1}(I) = \{x : \sqrt{y} = x, y \in I\}$.

On $[-1, 1]$, $f(x) = x^2$ has the image set $I = [0, 1]$. For any $y \neq 0$, $y \in I$, we can solve for $x = \pm\sqrt{y}$, so that x^2 is not invertible on $[-1, 1]$. However, the inverse image $f^{-1}([0, 1])$ is $[-1, 1]$, by the same reasoning.

There is an important generalization of a function called a *correspondence*:

Definition 2.2.5 *Let X and Y be sets. A* correspondence *is a rule that assigns a subset $F(x) \subseteq Y$ to each $x \in X$. Alternatively*[1]*, a correspondence is a rule that maps X into the power set of Y, $F : X \to 2^Y$.*

So a correspondence is just a "function with set-valued images": The restriction on functions that $f(x) = y$ and y is a singleton is being relaxed so that $F(x) = Z$ where Z is a set containing elements of Y.

Example Consider the equation $x^2 = c$. We can think of the correspondence $X(c)$ as the set of (real) solutions to the equation. For $c < 0$, the equation cannot be solved, because $x^2 > 0$, so $X(c)$ is empty for $c < 0$. For $c = 0$, there is exactly one solution, $X(c) = \{0\}$, and $X(c)$ happens to be a "function at that point". But for $c > 0$, $X(c) = \{\sqrt{c}, -\sqrt{c}\}$, since $(-1)^2 = 1$: Here there is a non-trivial correspondence, where the image is a set with multiple elements. ∎

Example Suppose two agents are playing "matching pennies". Their payoffs can be succinctly enumerated in a table:

			B
		H	T
	H	1,-1	-1,1
A	T	-1,1	1,-1

Suppose agent B uses a *mixed strategy* and plays randomly, so that $\sigma = pr$[Column uses H]. Then A's expected utility from using H is

$$\sigma 1 + (1 - \sigma)(-1)$$

while the row player's expected utility from using T is

$$\sigma(-1) + (1 - \sigma)(1)$$

We ask, "When is it an optimal strategy for the row player to use H against the column player's mixed strategy σ?", or, "What is the row player's *best-response* to σ?"

Well, H is strictly better than T if

$$\sigma 1 + (1 - \sigma)(-1) > \sigma(-1) + (1 - \sigma)(1) \to \sigma > \frac{1}{2}$$

[1]For reasons beyond our current purposes, there are advantages and disadvantages to each approach.

and H is strictly worse than T if

$$\sigma(-1) + (1-\sigma)(1) > \sigma 1 + (1-\sigma)(-1) \rightarrow \sigma < \frac{1}{2}$$

But when $\sigma = 1/2$, the row player is exactly indifferent between H and T. Indeed, the row player can himself randomize over any mix between H and T and get exactly the same payoff. Therefore, the row player's *best-response correspondence* is

$$pr[\text{Row uses H}|\sigma] = \begin{cases} 1 & \sigma > 1/2 \\ 0 & \sigma < 1/2 \\ [0,1] & \sigma = 1/2 \end{cases}$$

So correspondences arise naturally (and frequently!) in game theory. ∎

Example Consider the maximization problem

$$\max_{x_1, x_2} x_1 + x_2$$

subject to the constraints $x_1, x_2 \geq 0$, $x_1 + px_2 \leq 1$.

Since x_1 and x_2 are perfect substitutes, the solution is intuitively to pick the cheapest good and buy only that one. However, in the case when $p = 1$, the objective and constraint lie on top of one another, and any pair $x_1^* + x_2^* = 1$ is a solution. The solution to the maximization problem then is

$$(x_1^*(p), x_2^*(p)) = \begin{cases} (0, 1/p) & p < 1 \\ (1, 0) & p > 1 \\ (z, 1-z) & p = 1, z \in [0,1] \end{cases}$$

Therefore, we have a correspondence, where the optimal solution in the case when $p = 1$ is set valued, $[z, 1-z]$ for $z \in [0,1]$. So correspondences arise naturally in optimization theory. ∎

It turns out that correspondences are very common in microeconomics, even though they aren't usually studied in calculus or undergraduate math courses.

2.3 Real Numbers and Absolute Value

The set of numbers we usually work in, $(-\infty, \infty)$, is called the *real numbers*, or $\mathbb{R}$. The symbols $-\infty$ and ∞ are not in $\mathbb{R}$ (they are not even really numbers), but represent the idea of an unbounded process, like $1, 2, 3, ...$. The features of $\mathbb{R}$ that make it attractive follow from the intuitive idea that $\mathbb{R}$ is a continuum, with no breaks. This is unlike the integers, $\mathbb{Z} = ..., -3, -2, -1, 0, 1, 2, 3, ...$, which have spaces between each number.

Absolute value is given by

$$|x| = \begin{cases} -x & \text{if } x < 0 \\ 0 & \text{if } x = 0 \\ x & \text{if } x > 0 \end{cases}$$

The most important feature of absolute value is that it satisfies

$$|x + y| \leq |x| + |y|$$

And in particular,

$$|x - y| \leq |x| + |y|$$

2.4 Logic and Methods of Proof

It really helps to get a sense of what is going on in this section, and return to it a few times over the course. Thinking logically and finding the most *elegant* (read as, "slick") way to prove something is a skill that is developed. It is a talent, like composing or improvising music or athletic ability, in that you begin with a certain stock of potential to which you can add by working hard and being alert and careful when you see how other people do things. Even if you want to be an applied micro or macro economist, you will someday need to make logical arguments that go beyond taking a few derivatives or citing someone else's work, and it will be easier if you remember the basic nuts and bolts of what it means to "prove" something.

Propositional Logic

Definition 2.4.1 *A* proposition *is a statement that is either true or false.*

For example, some propositions are

- "The number e is transcendental"
- "Every Nash equilibrium is Pareto optimal"
- "Minimum wage laws cause unemployment"

These statements are either true or false. If you think minimum wage laws *sometimes* cause unemployment, then you are simply asserting the third proposition is false, although a similar statement that is more tentative might be true. The following are not propositions:

- "What time is it?"
- "This statement is false"

The first sentence is simply neither true nor false. The second sentence is such that if it were true it would be false, and if it were false it would be true[2]. Consequently, you can arrange symbols in ways that do not amount to propositions, so the definition is not meaningless.

Our goal in theory is usually to establish that "If P is true, then Q must be true", or "$P \to Q$", or " Q is a logical implication of P". We begin with a set of conditions that take the form of propositions, and show that these propositions collectively imply the truth of another proposition.

Since any proposition P is either true or false, we can consider the proposition's *negation*: The proposition that is true whenever P is false, and false whenever P is true. We write the negation of P in symbols as $\neg P$ as "not P". For example, "the number e is not transcendental", "some Nash equilibria are not Pareto optimal", and "minimum wage laws do not cause unemployment". Note that when negating a proposition, you have to take some care, but we'll get to more on that later.

Consider the following variations on "$P \to Q$":

- The Converse: $Q \to P$
- The Contrapositive: $\neg Q \to \neg P$
- The Inverse: $\neg P \to \neg Q$

[2]Note that this sentence is not a proposition not because it "contradicts" the definition of a proposition, but because it fails to satisfy the definition.

The converse is usually false, but sometimes true. For example, a differentiable function is continuous ("differentiable $\to$ continuous" is true), but there are continuous functions that are not differentiable ("continuous $\to$ differentiable" is false). The inverse, likewise, is usually false, since non-differentiable functions need not be discontinuous, since many functions with kinks are continuous but non-differentiable, like $|x|$. The contrapositive, however, is always true if $P \to Q$ is true, and always false if $P \to Q$ is false. For example, "a discontinuous function cannot be differentiable" is a true statement. Another way of saying this is that a claim $P \to Q$ and its contrapositive have the same truth value (true or false). Indeed, it is often easier to prove the contrapositive than to prove the original claim.

While the above paragraph shows that the converse and inverse need not be true if the original claim $P \to Q$ is true, we should show more concretely that the contrapositive is true. Why is this? Well, if $P \to Q$, then whenever P is true, Q must be true also, which we can represent in a table:

P	Q	P $\to Q$
T	T	T
T	F	F
F	T	T
F	F	T

So the first two columns say what happens when P and Q are true or false. If P and Q are both true, then of course $P \to Q$ is true. But if P and Q are both false, then of course $P \to Q$ is also a true statement. The proposition "$P \to Q$" is actually only false when P is true but Q is false, which is the second line. Now, as for the contrapositive, the truth table is:

P	Q	$\neg Q \to \neg P$
T	T	T
T	F	F
F	T	T
F	F	T

By the same kind of reasoning, when P and Q are both true or both false, $\neg Q \to \neg P$ is true. In the case where P is true but Q is false, we have $\neg Q$ implies $\neg P$, but P is true, so $\neg Q \to \neg P$ is false. Thus, we end up with exactly the same truth values for the contrapositive as for the original claim, so they are equivalent. To see whether you understand this, make a truth table for the inverse and converse, and compare them to the truth table for the original claim.

Quantifiers and negation

Once you begin manipulating the notation and ideas, you must inevitably negate a statement like, "All bears are brown." This proposition has a quantifier in it, namely the word "all". In this situation, we want to assert that "All bears are brown" is false; there are, after all, giant pandas that are black and white and not brown at all. So the right negation is something like, "Some bears are not brown". But how do we get "some" from "all" and "not brown" from "brown" in more general situations without making mistakes?

Definition 2.4.2 *Given a set A, the set of elements S that share some property P is written $S = \{x \in A : P(x)\}$ (e.g., $S_1 = \{x \in \mathbb{R} : 2x > 4\} = \{x \in \mathbb{R} : x \geq 2\}$) .*

Definition 2.4.3 *In a set A, if all the elements x share some property P, $\forall x \in A : P(x)$. If there exists an element $x \in A$ that satisfies a property P, then write $\exists x \in A : P(x)$. The symbol $\forall$ is the* universal quantifier *and the symbol $\exists$ is the* existential quantifier*.*

So a claim might be made,

- All allocations of goods in a competitive equilibria of an economy are Pareto optimal.

Then we are implicitly saying: Let A be the set of all allocations generated by a competitive equilibrium, and let $P(a)$ denote the claim that a is a "Pareto optimal allocation", whatever that is. Then we might say, "If $a \in A$, $P(a)$", or

- $\forall a \in A : P(a)$

Negating propositions with quantifiers can be very tricky. For example, consider the claim: "Everybody loves somebody sometime." We have three quantifiers, and it is unclear what order they should go in. The negation of this statement becomes a non-trivial problem precisely because of how many quantifiers are involved: Should the negation be, "Everyone hates someone sometime?" Or "Someone hates everyone all the time"? It requires care to get these details right. Recall the definition of negation:

Definition 2.4.4 *Given a statement $P(x)$, the* negation *of $P(x)$ asserts that $P(x)$ is false for x. The negation of a statement $P(x)$ is written $\neg P(x)$.*

The rules for negating statements are

$$\neg(\forall x \in A : P(x)) = \exists x \in A : \neg P(x)$$

and

$$\neg(\exists x \in A : P(x)) = \forall x \in A : \neg P(x)$$

For example, the claim "All allocations of goods in a competitive equilibria of an economy are Pareto optimal" could be written in the above form as follows: Let A be the set of allocations achievable in a competitive equilibrium of an economy (whatever that is), let $a \in A$ be a particular allocation, and let $P(a)$ be the proposition that the allocation is Pareto optimal (whatever that is). Then the claim is equivalent to the statement $\forall a \in A : P(a)$. Then the negation of that statement, then, is

- $\exists a \in A : \neg P(a)$

or in words, "There exists an allocation of goods in some competitive equilibrium of an economy that is not Pareto optimal."

Be careful about "or" statements. In proposition logic, $P \rightarrow Q_1$ or Q_2 means, "P logically implies Q_1, or Q_2, or both". So the negation of "or" statements is that "P implies either not Q_1 or not Q_2 or neither". For example, "All U.S. Presidents were U.S. citizens and older than 35 when they took office" is negated to "There's no U.S. President who wasn't a U.S. citizen, or younger than 35, or both, when they took office."

Some examples:

- "A given strategy profile is a Nash equilibrium if there is no player with a profitable deviation." Negating this statement gives, "A given strategy profile is not a Nash equilibrium if there exists a player with a profitable deviation."

- "A given allocation is Pareto efficient if there is no agent who can be made strictly better off without making some other agent worse off." Negating this statement gives, "A given allocation is not Pareto efficient if there exists an agent who can be made strictly better off without making any other agent worse off."

- "A continuous function with a closed, bounded subset of Euclidean space as its domain always achieves its maximum." Negating this statement gives, "A function that is not continuous or whose domain is not a closed, bounded subset of Euclidean space may not achieve a maximum."

Obviously, we are not logicians, so our relatively loose statements of ideas will have to be negated with some care.

2.4.1 Examples of Proofs

Most proofs use one of the following approaches:

- Direct proof: Show that P logically implies Q
- Proof by Contrapositive: Show that $\neg Q$ logically implies $\neg P$
- Proof by Contradiction: Assume P and $\neg Q$, and show this leads to a logical contradiction
- Proof by Induction: Suppose we want to show that for any natural number $n = 1, 2, 3, ...$, $P(n) \to Q(n)$. A proof by induction shows that $P(1) \to Q(1)$ (the *base case*), and that for any n, if $P(n) \to Q(n)$ (the *induction hypothesis*), then $P(n+1) \to Q(n+1)$. Consequently, $P(n) \to Q(n)$ is true for all n.
- "Disproof by Counterexample": Let $\forall x \in A : P(x)$ be a statement we wish to disprove. Show that $\exists x' \in A$ such that $P(x')$ is false.

The claim might be fairly complicated, like

- If "$f(x)$ is a bounded, increasing function on $[a, b]$", then "the set of points of discontinuity of $f()$ is a countable set".

The "P" is, ($f(x)$ is a bounded, increasing function on $[a, b]$), and the "Q" is (the set of points of discontinuity of $f(x)$ is a countable set). To prove this, we could start by using P to show Q must be true. Or, in a proof by contrapositive, we could prove that if the function has an uncountable number of discontinuities on $[a, b]$, then $f(x)$ is either unbounded, or decreasing, or both. Or, in a proof by contradiction, we could assume that the function is bounded and increasing, but that the function has an uncountable number of discontinuities. The challenge for us is to make sure that while we are using mathematical sentences and definitions, no logical mistakes are made in our arguments.

Each method of proof has its own advantages and disadvantages, so we'll do an example of each kind now.

A Proof by Contradiction

Theorem 2.4.5 *There is no largest prime.*

Proof Suppose there was a largest prime number, n. Let $p = n * (n-1) * ... * 1 + 1$. Then p is not divisible by any of the first n numbers, so it is prime and greater than n, since $n*(n-1)*..*1+1 > n*1+1 > n$. So p is a prime number larger than n — this is a contradiction, since n was assumed to be the largest prime. Therefore, there is no largest prime. ∎

A Proof by Contrapositive

Suppose we have two disjoint sets, M and W with $|M| = |W|$, and are looking for a *matching* of elements in M to elements in W — a relation mapping each element from M into W and vice versa, which is one-to-one and onto. Each element has a "quality" attached, either $q_m \geq 0$ or $q_w \geq 0$, and the value of the match is $v(q_m, q_w) = \frac{1}{2} q_m q_w$. The *assortative match* is the one where the highest quality agents are matched, the second-highest quality agents are matched, and so on. If agent m and agent w are matched, then $h(m, w) = 1$; otherwise $h(m, w) = 0$.

Theorem 2.4.6 *If a matching maximizes* $\sum_m \sum_w h(m, w) v(q_m, q_w)$, *then it is assortative.*

Proof Since the proof is by contrapositive, we need to show that any matching that is not assortative does not maximize $\sum_m \sum_w h_{mw} v(q_m, q_w)$ — i.e., any non-assortative match can be improved.

If the match is not assortative, we can find two sets of agents where $q_{m_1} > q_{m_2}$ and $q_{w_1} > q_{w_2}$, but $h(m_1, w_2) = 1$ and $h(m_2, q_1) = 1$. Then the value of those two matches is

$$q_{m_1} q_{w_2} + q_{m_2} q_{w_1}$$

If we swapped the partners, the value would be

$$q_{m_1} q_{w_1} + q_{m_2} q_{w_2}$$

and

$$\{q_{m_1} q_{w_1} + q_{m_2} q_{w_2}\} - \{q_{m_1} q_{w_2} + q_{m_2} q_{w_1}\} = (q_{m_1} - q_{m_2})(q_{w_1} - q_{w_2}) > 0$$

So a non-assortative match does not maximize $\sum_m \sum_w h(m, w) v(q_m, q_w)$. $\blacksquare$.

What's the difference between proof by contradiction and proof by contrapositive? In proof by contradiction, you get to use P and $\neg Q$ to arrive at a contradiction, while in proof by contrapositive, you use $\neg Q$ to show that $\neg P$ is a logical implication, without using P or $\neg P$ in the proof.

A Proof by Induction

Theorem 2.4.7 $\sum_{i=1}^{n} i = \frac{1}{2} n(n+1)$

Proof *Basis Step:* Let $n = 1$. Then $\sum_{i=1}^{1} i = 1 = \frac{1}{2} 1(1+1) = 1$. So the theorem is true for $n = 1$.

Induction Step: Suppose $\sum_{i=1}^{n} i = \frac{1}{2} n(n+1)$; we'll show that $\sum_{i=1}^{n+1} i = \frac{1}{2}(n+1)(n+2)$. Now add $n+1$ to both sides:

$$\sum_{i=1}^{n} i + (n+1) = \frac{1}{2} n(n+1) + (n+1)$$

Factoring $\frac{1}{2}(n+1)$ out of both terms gives

$$\sum_{i=1}^{n+1} i = \frac{1}{2}(n+1)(n+2)$$

And the induction step is shown to be true.

Now, since the basis step is true ($P(1) \to Q(1)$ is true), and the induction step is true ("$P(n) \to Q(n)$" $\to$ "$P(n+1) \to Q(n+1)$"), the theorem is true for all n. $\blacksquare$

Disproof by Counter-example

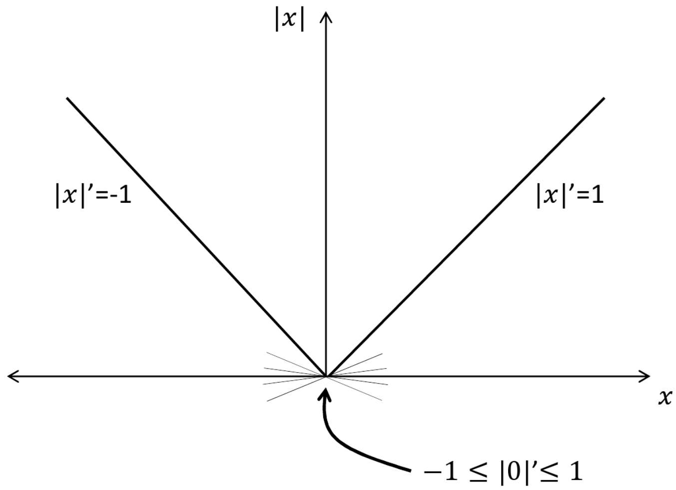
Definition 2.4.8 *Given a (possibly false) statement $\forall x \in A : P(x)$, a* counter-example *is an element x' of A for which $\neg P(x')$.*

For example, "All continuous functions are differentiable." This sounds like it could be true. Continuous functions have no jumps and are pretty well-behaved. But it's false. The negation of the statement is, "There exist continuous functions that are not differentiable." Maybe we can prove that?

My favorite non-differentiable function is the absolute value,

$$|x| = \begin{cases} -x & \text{if } x < 0 \\ 0 & \text{if } x = 0 \\ x & \text{if } x > 0 \end{cases}$$

For $x < 0$, the derivative is $f'(x) = -1$, and for $x > 0$, the derivative is $f'(x) = 1$. But *at* zero, the derivative is not well-defined. Is it $+1$? -1? Something in-between?



Non-uniqueness of $|0|'$

This is what we call a *kink* in the function, a point of non-differentiability. So $|x|$ is non-differentiable at zero. But $|x|$ is continuous at zero, since for any sequence $x_n \to 0$, $|x_n| \to |0|$. So we have provided a *counterexample*, showing that "All continuous functions are differentiable" is false.

One theorem three ways

We did three theorems that highlight the usefulness of each method of proof above. Here, we'll do one theorem using each of the three tools. Our theorem is[3]:

Theorem 2.4.9 *Suppose $f(x)$ is continuously differentiable. If $f(x)$ is increasing on $[a, b]$, then $f'(x) \geq 0$ for all $x \in [a, b]$.*

Now, the direct proof:

Proof Let $x' > x$. If $f(x)$ is increasing, then $f(x') > f(x)$, and by the fundamental theorem of calculus,

$$f(x') - f(x) = \int_x^{x'} f'(z)dz$$

Since x' and x are arbitrary, let $x' = x + \varepsilon$, where $\varepsilon > 0$. Then

$$\frac{f(x+\varepsilon) - f(x)}{\varepsilon} = \frac{\int_x^{x+\varepsilon} f'(z)dz}{\varepsilon} > 0$$

and taking the limit as $\varepsilon \downarrow 0$ gives, $f'(x) > 0$.

[3]A continuously differentiable function is one for which its derivative exists everywhere, and is itself a continuous function.

In this proof, we use the premise to prove the conclusion directly, without using the conclusion. Let's try a proof by contrapositive.

Proof This proof is by contrapositive, so we will show that if $f'(x) < 0$ for all $x \in [a, b]$, then $f(x)$ is decreasing.

Suppose $f'(x) < 0$ for all $x \in [a, b]$, and $x' > x$ are in $[a, b]$. Then

$$f(x') - f(x) = \int_{x}^{x'} f'(z)dz < 0$$

so that $f(x') < f(x)$, as was to be shown.

Here, instead of showing that an increasing function has positive derivatives, we showed that negative derivatives implied a decreasing function. In some sense, the proof was easier, even though the approach involved more thought. The last proof is by contradiction:

Proof This proof is by contradiction, so we assume that for all $x' > x$ in $[a, b]$, $f(x') > f(x)$, but $f'(x) < 0$ for all $x \in [a, b]$.

If $x' > x$, then $f(x') - f(x) = \int_{x}^{x'} f'(z)dz < 0$, so that $f(x') < f(x)$. But we assumed that $f(x') > f(x)$, so we are lead to a contradiction.

See how we get to use both $\not P$ and Q to arrive at a contradiction? That is the key advantage. The disadvantage is that you can often learn more from a direct proof about the mathematical situation than you can from a proof by contradiction, since the first often constructs the objects of interest or shows how various conditions give rise to certain properties.

Necessary and Sufficient Conditions

If $P \to Q$ and $Q \to P$, then we often say "P *if and only if* Q", or "P is *necessary* for Q and Q is *sufficient* for P".

Consider

- To be a lawyer, it is necessary to have attended law school.
- To be a lawyer, it is sufficient to have passed the bar exam.

To pass the bar exam, it is required that a candidate have a law degree. On the other hand, a law degree alone does not entitle someone to be a lawyer, since they still need to be certified. So a law degree is a necessary condition to be a lawyer, but not sufficient. On the other hand, if someone has passed the bar exam, they are allowed to represent clients, so it is a sufficient condition to be a lawyer. There is clearly a gap between the necessary condition of attending law school, and the sufficient condition of passing the bar exam; namely, passing the bar exam.

Let's look at a more mathematical example:

- A necessary condition for a point x^* to be a local maximum of a differentiable function $f(x)$ on (a, b) is that $f'(x^*) = 0$.
- A sufficient condition for a point x^* to be a local maximum of a differentiable function $f(x)$ on (a, b) is that $f'(x^*) = 0$ and $f''(x^*) < 0$.

Why is the first condition necessary? Well, if $f(x)$ is differentiable on (a, b) and $f'(x^*) > 0$, we can take a small step to $x^* + h$ and increase the value of the function (alternatively, if $f'(x^*) < 0$, take a small step to $x^* - h$ and you can also increase the value of the function). Consequently, x^* could not have been a local maximum. So it is necessary that $f'(x^*) = 0$.

Why is the second condition sufficient? Well, Taylor's theorem says that

$$f(x) = f(x^*) + f'(x^*)(x - x^*) + f''(x^*)\frac{(x^* - x)^2}{2} + o((h)^3)$$

where $x^* - x = h$ and $o(h^3)$ goes to zero as h goes to zero faster than the $(x^* - x)^2$ and $(x^* - x)$ terms. Since $f'(x^*) = 0$, we then have

$$f(x) = f(x^*) + f''(x^*)\frac{(x - x^*)^2}{2} + o(h^3)$$

so that for x sufficiently close to x^*,

$$f(x^*) - f(x) = -f''(x^*)\frac{(x^* - x)^2}{2}$$

so that $f(x^*) > f(x)$ only if $f''(x^*) < 0$. So if the first-order necessary conditions and second-order sufficient conditions are satisfied, x^* must be a solution. However, there is a substantial gap between the first-order necessary and second-order sufficient conditions.

For example, consider $f(x) = x^3$ on $(-1, 1)$. The point $x^* = 0$ is clearly a solution of the first-order necessary conditions, and $f''(0) = 0$ as well. However, $f(1/2) = (1/2)^3 > 0 = f(0)$, so the point x^* is a critical point, but not a maximum. Similarly, suppose we try to maximize $f(x) = x^2$ on $(-1, 1)$. Again, $x^* = 0$ is a solution to the first-order necessary conditions, but it is a *minimum*, not a maximum. So the first-order necessary conditions only identify critical points, but these critical points come in three flavors: maximum, minimum, and saddle-point.

Alternatively, consider the function

$$g(x) = \begin{cases} x & , -\infty < x < 5 \\ 5 & , 5 \le x \le 10 \\ 10 - x & , 10 < x < \infty \end{cases}$$

This function looks like a plateau, where the set $[5, 10]$ all sits at a level of 5, and everything before and after drops off. The necessary conditions are satisfied for all x strictly between 5 and 10, since $f'(x) = 0$ for x strictly between 5 and 10. But the sufficient conditions aren't satisfied, since $f''(x) = 0$ for x strictly between 5 and 10. These points are all clearly global maxima, but the second-order sufficient conditions require *strict* negativity to rule out pathologies like $f(x) = x^3$.

So, even when you have useful necessary and sufficient conditions, there can be subtle gaps that you miss if you're not careful.

Exercises

1. If A and B are sets, prove that $A \subseteq B$ iff $A \cap B = A$.

2. Prove the distributive law

$$A \cap (B \cap C) = (A \cap B) \cap (A \cap C)$$

and the associative law

$$(A \cup B) \cup C = A \cup (B \cup C)$$

What is $((A \cup B) \cup))^c$?

3. Sketch graphs of the following sets:

$$\{(x, y) \in \mathbb{R}^2 : x \ge y\}$$

$$\{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq 1\}$$
$$\{(x, y) \in \mathbb{R}^2 : \sqrt{xy} \geq 5\}$$
$$\{(x, y) \in \mathbb{R}^2 : \min\{x, y\} \geq 5\}$$

4. Write the converse and contrapositive of the following statements (note that you don't have to know what any of the statements actually mean to negate them) :

- A set X is *convex* if, for every $x_1, x_2 \in X$ and $\lambda \in [0, 1]$, the element $x_\lambda = \lambda x_1 + (1 - \lambda)x_2$ is in X.
- The point x^* is a maximum of a function f on the set A if there exists no $x' \in A$ such that $f(x') > f(x)$.
- A strategy profile is a *subgame perfect Nash equilibrium* if, for every player and in every subgame, the strategy profile is a Nash equilibrium.
- A function is *uniformly continuous on a set* A if, for all $\varepsilon > 0$ there exists a $\delta > 0$ such that for all $x, c \in A$, $|x - c| < \delta$, $|f(x) - f(c)| < \varepsilon$.
- Any competitive equilibrium (x^*, y^*, p^*) is a Pareto optimal allocation.

Part I

Optimization and Comparative Statics in $\mathbb{R}$

Chapter 3

Basics of $\mathbb{R}$

Deliberate behavior is central to every microeconomics model, since we assume that agents act to get the best payoff possible available to them, given the behavior of other agents and exogenous circumstances that are outside their control.

Some examples are:

- A firm hires capital K and labor L at prices r and w per unit, respectively, and has production technology function $F(K, L)$. It receives a price p per unit it sells of its good. It would like to maximize profits,
$$\pi(K, L) = pF(K, L) - rK - wL$$
- A consumer has utility function $u(x)$ over bundles of goods $x = (x_1, x_2, ..., x_N)$, and a budget constraint $w \leq \sum_{i=1}^N p_i x_i$. The consumer wants to maximize utility, subject to his budget constraint:
$$\max_x u(x)$$
subject to $\sum_i p_i x_i \leq w$.
- A buyer goes to a first-price auction, where the highest bidder wins the good and pays his bid. His value for the good is v, and the probability he wins, conditional on his bid, is $p(b)$. If he wins, he gets a payoff $v - b$, while if he loses, he gets nothing. Then he is trying to maximize
$$\max_b p(b)(v - b) + (1 - p(b))0$$

These are probably very familiar to you. But, there are other, related problems of interest you might not have thought of:

- A firm hires capital K and labor L at prices r and w per unit, respectively, and has production technology function $F(K, L)$. It would like to minimize costs, subject to producing $\bar{q}$ units of output, or
$$\min_{K,L} rK + wL$$
subject to $F(K, L) = \bar{q}$.
- A consumer has utility function $u(x)$ over bundles of goods $x = (x_1, x_2, ..., x_N)$, and a budget constraint $w \leq \sum_{i=1}^N p_i x_i$. The consumer would like to minimize the amount they spend, subject to achieving utility equal to $\bar{u}$, or
$$\min_x \sum_{i=1}^N p_i x_i$$
subject to $u(x) \geq \bar{u}$.

- A seller faces N buyers, each of whom has a value for the good unobservable to the seller. How much revenue can the seller raise from the transaction, among all possible ways of soliciting information from the buyers?

We want to develop some basic, useful facts about solving these kinds of models. There are three questions we want to answer:

- When do solutions exist to maximization problems? (The Weierstrass theorem)
- How do we find candidate solutions? (First-order necessary conditions)
- How do we show that a candidate solution is a maximum? (Second-order sufficient conditions)

Since we're going to study dozens of different maximization problems, we need a general way of expressing the idea of a maximization problem efficiently:

Definition 3.0.10 *A* maximization problem *is a* payoff function *or* objective function*, $f(x,t)$, and a* choice set *or* feasible set*, X. The agent is then trying to solve*

$$\max_{x \in X} f(x,t)$$

The variable x is a choice variable *or* control*, since the agent gets to decide its value. The variable t is an* exogenous variable *or* parameter*, over which the agent has no control.*

For this chapter, agents can choose any real number: Firms pick a price, consumers pick a quantity, and so on. This means that X is the set of real numbers, $\mathbb{R}$. Also, unless we are interested in t in particular, we often ignore the exogenous variables in an agent's decision problem. For example, we might write $\pi(q) = pq - cq^2$, even though $\pi()$ depends on p and c.

A good question to start with is, "When does a maximizer of $f(x)$ exist?" While this question is simple to state in words, it is actually pretty complicated to state and show precisely, which is why this chapter is fairly "math-heavy".

3.0.2 Maximization and Minimization

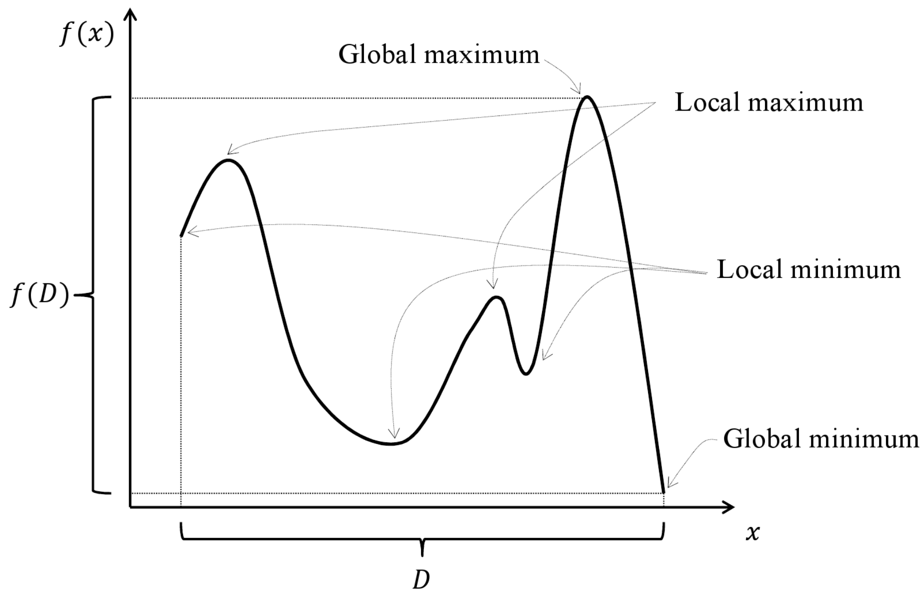
A *function* is a mapping from some set D, the *domain*, uniquely into $\mathbb{R} = (-\infty, \infty)$, *the real numbers*, and we write $f : D \to \mathbb{R}$. For example, $\sqrt{x} : [0, \infty) \to \mathbb{R}$. The *image of D under $f(x)$* is the set $f(D)$. For example, $\sqrt{D} = [0, \infty)$, and for $\log(x) : (0, \infty) \to \mathbb{R}$, we have $\log(D) = (-\infty, \infty)$. It turns out that the properties of the domain D and the function $f(x)$ both matter for us.

Definition 3.0.11 *A point x^* is a* global maximizer of $f : D \to \mathbb{R}$ *if, for any other x' in D,*

$$f(x^*) \geq f(x')$$

A point x^* is a local maximizer of $f : D \to \mathbb{R}$ *if, for some $\delta > 0$ and for any other x' in $(x^* - \delta, x^* + \delta) \cap D$,*

$$f(x^*) \geq f(x')$$



Domain, Image, Maxima, Minima

So a local maximum is basically saying, "If we restrict attention to $(x^* - \delta, x^* + \delta)$ instead of (a, b), then x^* is a global maximum in $(x^* - \delta, x^* + \delta)$." In even less mathematical terms, "x^* is a local maximum if it does at least as well as anything nearby."

A point $f(x^*)$ is a local or global *minimum* if, instead of $f(x^*) \geq f(x')$ above, we have $f(x^*) \leq f(x')$. However, we can focus on just maximization or just minimization, since

Theorem 3.0.12 *If x^* is a local maximum of $f()$, then x^* is a local minimum of $-f()$.*

Proof If $f(x^*) \geq f(x')$ for all x' in $(x^* - \delta, x^* + \delta)$ for some $\delta > 0$, then $-f(x^*) \leq -f(x')$ for all x' in $(x^* - \delta, x^* + \delta)$, so it is a local minimum. ∎

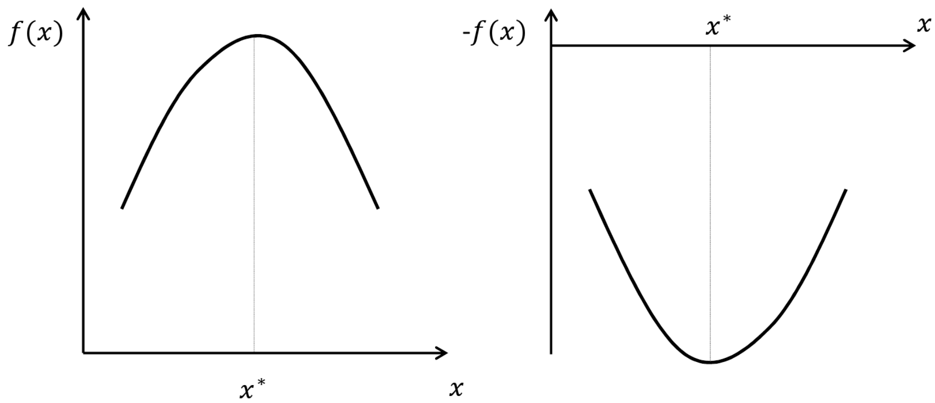
So whenever you have a minimization problem,

$$\min_x f(x)$$

you can turn it into a maximization problem

$$\max_x -f(x)$$

and forget about minimization entirely.



Maximization vs. Minimization

On the other hand, since I have programmed myself to do this automatically, I find it annoying when a book is written entirely in terms of minimization, and some of the rules used down the road for minimization (second-order sufficient conditions for constrained optimization problems) are slightly different than those for maximization.

3.1 Existence of Maxima/Maximizers

So what do we mean by "existence of a maximum or maximizer"?

Before moving to functions, let's start with the maximum and minimum of a set. For a set $F = [a, b]$ with a and b finite, the maximum is b and the minimum is a — these are members of the set, and $b \geq x$ for all other x in S, and $a \leq x$ for all other x in S.

If we consider the set $E = (a, b)$ however, the points b and a are not actually members of the set, although $b \geq x$ for all x in E, and $a \leq x$ for all x in E. We call b the *least upper bound* of E, or the *supremum*,

$$\sup(a, b) = b$$

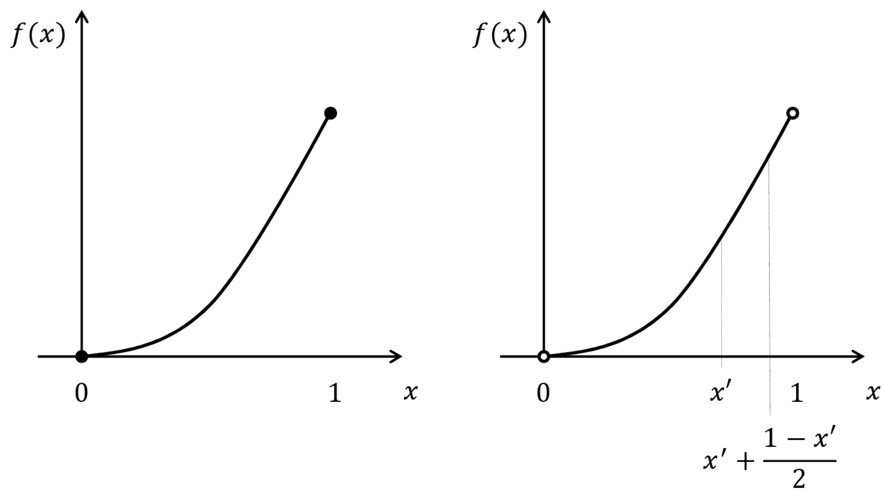
and a is the *greatest lower bound* of E, or the *infimum*,

$$\inf(a, b) = a$$

Even if the maximum and minimum of a set don't exist, the supremum and infimum always do (though they may be infinite, like for (a, ∞)).

Since function is a rule that assigns a real number to each x in the domain D, the image of D under f is going to be a set, $f(D)$. For maximization purposes, we are interested in whether the image of a function includes its largest point or not.

For example, the function $f(x) = x^2$ on $[0, 1]$ achieves a maximum at $x^* = 1$, since $f(1) = 1 > f(x)$ for $0 \leq x < 1$, and $\sup f([0, 1]) = \max f([0, 1])$. However, if we consider $f(x) = x^2$ on $(0, 1)$, the supremum of $f(x)$ is still 1, since 1 is the least upper bound of $f((0, 1)) = (0, 1)$, but the function does not achieve a maximum.



Existence and Non-existence

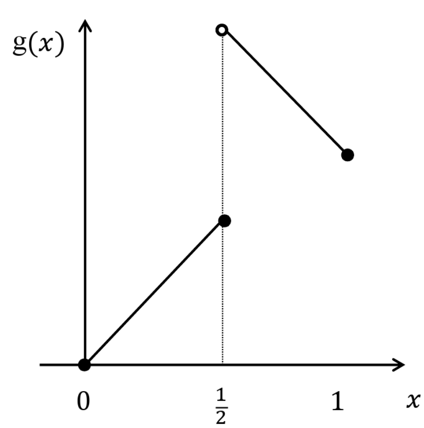
Why? On the left, the graph has no "holes", so it moves smoothly to the highest point. On the right, however, a point is "missing" from the graph exactly where it would have achieved its maximum. This is the difference between a maximizer (left) and a supremum when a maximizer fails to exist (right). Why does this happen? In short, because the choice set on the left, $[a, b]$, "includes all its points", while the choice set on the right, (a, b), is missing a and b.

To be a bit more formal, suppose $x' \in (0,1)$ is the maximum of $f(x) = x^2$ on $(0, 1)$. It must be less than 1, since 1 is not an element of the choice set. Let $x'' = (1 - x')/2 + x'$ — this is the point exactly halfway between x' and 1. Then $f(x'') > f(x')$, since $f(x)$ is increasing. But then x' couldn't have been a maximizer. This is true for all x' in $(0, 1)$, so we're forced to conclude that no maximum exists.

But this isn't the only way something can go wrong. Consider the function on $[0, 1]$ where

$$g(x) = \begin{cases} x & 0 \leq x \leq \dfrac{1}{2} \\ 2 - x & \dfrac{1}{2} < x \leq 1 \end{cases}$$

Again, the function's graph is missing a point precisely where it would have achieved a maximum, at 1/2. Instead, it takes the value 1/2 instead of 3/2.



So in either case, the function can fail to achieve a maximum. In one situation, it happens because the domain D has the wrong features: It is missing some key points. In the second situation, it happens because the function f has the wrong features: It jumps around in unfortunate ways.

Consequently, the question is: What features of the domain D and the function $f(x)$ guarantee that $\sup_{x \in D} f(x) = m^*$ is equal to $f(x^*)$ for some $x \in D$?

Definition 3.1.1 *A (global) maximizer of $f(x)$, x^*,* exists *if $f(x^*) = \sup f(x)$, and $f(x^*)$ is the* (global) maximum *of $f()$.*

3.2 Intervals and Sequences

Intervals are special subsets of the real numbers, but some of their properties are key for optimization theory.

Definition 3.2.1 *The set $(a, b) = \{x : a < x < b\}$ is* open, *while the set $[a, b] = \{x : a \leq x \leq b\}$ is* closed. *If $-\infty < a < b < \infty$, the sets (a, b) and $[a, b]$ are* bounded *(both the endpoints a and b are finite).*

A sequence $x_n = x_1, x_2, ...$ is a rule that assigns a real number to each of the *natural (counting) numbers*, $\mathbb{N} = 1, 2, 3, ...$

Some basic sequences are:

- Let $x_n = n$. Then the sequence is $1, 2, 3, ...$
- Let $x_n = \dfrac{1}{n}$. Then the sequence is $1, \dfrac{1}{2}, \dfrac{1}{3}, ...$
- Let $x_n = (-1)^{n-1}\dfrac{1}{n}$. Then the sequence is $1, -\dfrac{1}{2}, \dfrac{1}{3}, -\dfrac{1}{4} ...$
- Let $x_1 = 1$, $x_2 = 1$, and for $n > 2$, $x_n = x_{n-1} + x_{n-2}$. Then the sequence is $1, 1, 2, 3, 5, 8, 13, 21, ...$
- Let $x_n = \sin\left(\dfrac{n\pi}{2}\right)$. Then the sequence is $1, 0, -1, 0, ...$

At first, they might seem like a weird object to study when our goal is to understand functions. A function is basically an uncountable number of points, while sequences are countable, so it seems like we would lose information by simplifying an object this way. However, it turns out that studying how functions $f(x)$ act on a sequence x_n gives us all the information we need to study the behavior of a function, without the hassle of dealing with the function itself.

Definition 3.2.2 *A* sequence *$x_n = x_1, x_2, ...$ is a function from $\mathbb{N} \to \mathbb{R}$. A sequence* converges to *$\bar{x}$ if, for all $\varepsilon > 0$ there exists an N such that $n > N$ implies $|x_n - \bar{x}| < \varepsilon$. Then we write*

$$\lim_{n \to \infty} x_n = \bar{x}$$

and $\bar{x}$ is the limit *of x_n.*

Everyone's favorite convergent sequence is

$$x_n = \frac{1}{n} = 1, \frac{1}{2}, \frac{1}{3}, ...$$

and the limit of x_n is zero. Why? First, for any positive number ε, I can make $|x_n - 0| < \varepsilon$ by picking n large enough. Just take $n > N = 1/\varepsilon$, so that

$$\frac{1}{n} < \frac{1}{1/\varepsilon} = \varepsilon$$

Second, since ε was arbitrary, it might as well be zero. Therefore, $x_n \to 0$ as $n \to \infty$. Notice, however, how $1/n$ is never actually equal to zero for finite n; it is only in the limit that it hits its lower bound. This is another example of the difference between minimum and infimum: $\inf x_n = 0$ but $\min x_n$ does not exist, since $1/n > 0$ for all n.

Everyone's favorite non-convergent or *divergent* sequence is

$$x_n = n$$

Pick any finite number ε, and I can make $|x_n| > \varepsilon$, contrary to the definition of convergence. How? Just make $|n| > \varepsilon$. This sequence goes off to infinity, and its long-run behavior doesn't "settle down" to any particular number.

Definition 3.2.3 *A sequence x_n is bounded if $|x_n| \leq M$ for some finite real number M.*

But boundedness alone doesn't guarantee that a sequence converges, and just because a sequence fails to converge doesn't mean that it's not interesting. Consider

$$x_n = \cos\left(\frac{(n-1)\pi}{2}\right) = 1, 0, -1, 0, 1, 0, -1, ...$$

This function fails to converge (right?). But it's still quite well-behaved. Every even term is zero,

$$x_{2k} = 0, 0, 0$$

and it includes a convergent string of ones,

$$x_{4(k-1)+1} = 1, 1, 1, ...$$

and a convergent string of -1's,

$$x_{4(k-1)+3} = -1, -1, -1, ...$$

So the sequence x_n — even though it is non-convergent — is composed of three convergent sequences. These building blocks are important enough to deserve a name:

Definition 3.2.4 *Let n_k be a sequence of natural numbers, with $n_1 < n_2 < ... < n_k < ....$ Then x_{n_k} is a* sub-sequence *of x_n.*

It turns out that *every* bounded sequence has at least one sub-sequence that *is* convergent.

Theorem 3.2.5 (Bolzano-Weierstrass Theorem) *Every bounded sequence has a convergent subsequence.*

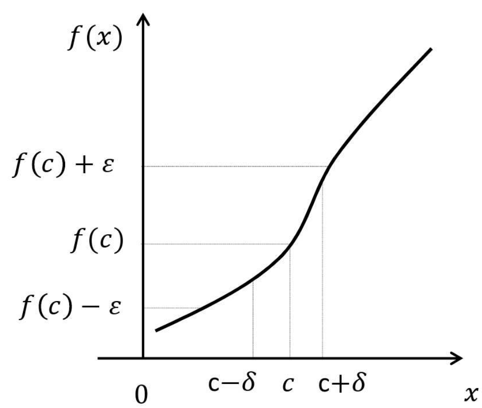
This seems like an esoteric (and potentially useless) result, but it is actually a fundamental tool in analysis. Many times, we have a sequence of interest but know very little about it. To study its behavior, we use the Bolzano-Weierstrass to find a convergent sub-sequence, and sometimes that's enough.

3.3 Continuity

In short, a function is continuous if you can trace its graph without lifting your pencil off the paper. This definition is not precise enough to use, but it captures the basic idea: The function has no "jumps" where the value $f(x)$ is changing very quickly even though x is changing very little. The reason continuity is so important is because continuous functions preserve properties of sets like being open, closed, or bounded. For example, if you take a set $S = (a, b)$ and apply a continuous function $f()$ to it, the *image* $f(S)$ is an open interval (c, d). This is not the case for every function.

The most basic definition of continuity is

Definition 3.3.1 *A function* $f(x)$ is continuous at c *if, for all* $\varepsilon > 0$*, there is a* $\delta > 0$ *so that if* $|x - c| < \delta$*, then* $|f(x) - f(c)| < \varepsilon$.



Continuity

Cauchy's original explanation was that ε is an "error tolerance", and a function is continuous at c if the actual error ($|f(x) - f(c)|$) can made smaller than any error tolerance by making x close to c ($|x - c| < \delta$). An equivalent definition of continuity is given in terms of sequences by the following:

Theorem 3.3.2 (Sequential Criterion for Continuity) *A function* $f(x)$ *is continuous at* c *iff for all sequences* $x_n \to c$*, we have*

$$\lim_{n\to\infty} f(x_n) = f\left(\lim_{n\to\infty} x_n\right) = f(c)$$

We could just as easily use "If $\lim_{n\to\infty} f(x_n) = f(c)$ for all sequences $x_n \to c$, then $f()$ is continuous" as the definition of continuity as the $\varepsilon - \delta$ definition. The Sequential Criterion for Continuity converts an idea about functions expressed into $\varepsilon - \delta$ notation into an equivalent idea about how $f()$ acts on sequences. In particular, a continuous function preserves convergence of a sequence x_n to its limit c.

The important result is that if $f(x)$ is continuous and $x_n \to c$, then

$$\lim_{n\to\infty} f(x_n) = f\left(\lim_{n\to\infty} x_n\right)$$

or that the limit operator and the function can be interchanged if and only if $f(x)$ is continuous.

3.4 The Extreme Value Theorem

Up to this point, what are the facts that you should internalize?

- Sequences are convergent only if they settle down to a constant value in the long run
- Every bounded sequence has a convergent subsequence, even if the original sequence doesn't converge (Bolzano-Weierstrass Theorem)
- A function is continuous if and only if $\lim_{n\to\infty} f(x_n) = f(x)$ for all sequences $x_n \to x$

With these facts, we can prove the *Weierstrass Theorem*, also called the *Extreme Value Theorem.*

Theorem 3.4.1 (Weierstrass' Extreme Value Theorem) *If $f : [a, b] \to \mathbb{R}$ is a continuous function and a and b are finite, then f achieves a global maximum on $[a, b]$.*

First, let's check that if any of the assumptions are violated, then examples exist where f does not achieve a maximum. Recall our examples of functions that failed to achieve maxima, $f(x) = x^2$ on $(0, 1)$ and

$$g(x) = \begin{cases} x & 0 \leq x \leq \dfrac{1}{2} \\ 2 - x & \dfrac{1}{2} < x \leq 1 \end{cases}$$

on $[0, 1]$. In the first example, $f(x)$ is continuous, but the set $(0, 1)$ is open, unlike $[a, b]$, violating the hypotheses of the Weierstrass theorem. In the second example, the function's domain is closed and bounded, but the function is discontinuous, violating the hypotheses of the Weierstrass theorem.

So the reason the Weierstrass theorem is useful is that it provides *sufficient* conditions for a function to achieve a maximum, so that we know for sure, without exception, that any continuous $f(x)$ on a closed, bounded set $[a, b]$ will achieve a maximum.

Proof Suppose that

$$\sup f(x) = m^*$$

We know $m^* < \infty$, since if $f(x)$ is continuous on $[a, b]$, it is well-defined for its whole domain and has no asymptotes (unlike $1/x$ on $[0, 1]$, which is discontinuous at zero). We want to show that there exists an x^* in $[a, b]$ so that $f(x^*) = m^*$.

Step 1: Since m^* is the supremum of $f()$, we can find a sequence of points $\{x_n\} = x_1, x_2, ...$ in $[a, b]$ so that

$$m^* - \frac{1}{n} \leq f(x_n) \leq m^*$$

Why? We show this by contradiction. If the above inequality was violated, then for some n, we would not be able to find an x_n so that $m^* - \dfrac{1}{n} \leq f(x_n) \leq m^*$, and $m^* - 1/n$ would be greater than $f(x)$ for all x in $[a, b]$, implying that $m^* - 1/n$ is the supremum of f AND less than m^*, so that m^* was not actually the supremum in the first place, since the supremum is the *least* upper bound. This would be a contradiction.

Step 2: Since the sequence x_n defined by $f(x_n)$ is contained in $[a, b]$, by the Bolzano-Weierstrass theorem we can find a convergent subsequence, $x_{n_k} \to x^*$, where x^* is in $[a, b]$.

Step 3: If we take the limit of the convergent subsequence,

$$\lim_{n_k\to\infty} m^* - \frac{1}{n_k} \leq \lim_{n_k\to\infty} f(x_{n_k}) \leq m^*$$

by continuity of $f()$ and the Sequential Criterion for Continuity, we have $\lim_{n_k\to\infty} f(x_{n_k}) = f(x^*)$, and

$$m^* \leq f(x^*) \leq m^*$$

The above inequalities imply $f(x^*) = m^*$, so a maximizer x^* exists in $[a, b]$, and the function achieves a maximum at $f(x^*)$. $\blacksquare$

This is the foundational result of all optimization theory, and it pays to appreciate how and why these steps are each required. This is the kind of powerful result you can prove using a little bit of analysis.

3.5 Derivatives

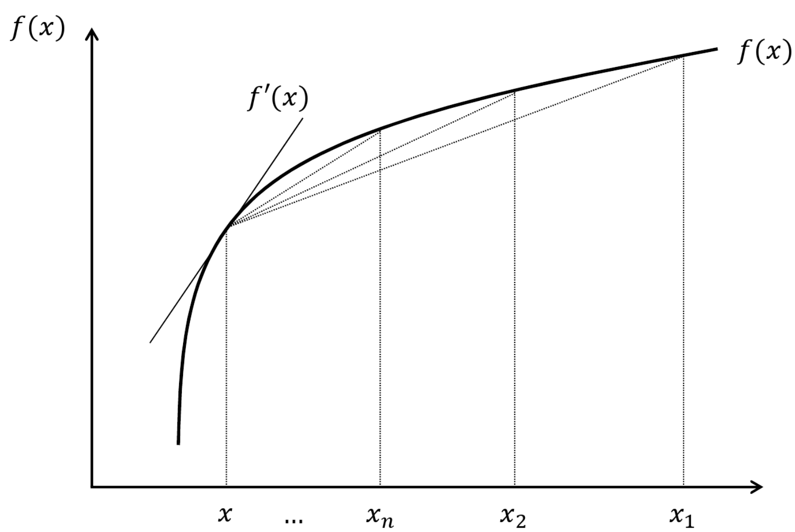
As you know, derivatives measure the rate of change of a function at a point, or

$$f'(x) = D_x f(x) = \frac{df(x)}{dx} = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h}$$

The way to visualize the derivative is as the limit of a sequence of chords,

$$\lim_{n \to \infty} \frac{f(x_n) - f(x)}{x_n - x}$$

that converge to the tangent line, $f'(x)$.



The Derivative

Since this sequence of chords is just a sequence of numbers, the derivative is just the limit of a particular kind of sequence. So if the derivative exists, it is unique, and a derivative exists only if it takes the same value no matter what sequence $x_n \to x$ you pick.

For example, the derivative of the square root of x, $\sqrt{x}$, can be computed "bare-handed" as

$$\lim_{h \to 0} \frac{\sqrt{x+h} - \sqrt{x}}{h} = \lim_{h \to 0} \frac{\sqrt{x+h} - \sqrt{x}}{h} \frac{\sqrt{x+h} + \sqrt{x}}{\sqrt{x+h} + \sqrt{x}}$$
$$= \lim_{h \to 0} \frac{x+h-x}{h} \frac{1}{\sqrt{x+h} + \sqrt{x}} = \lim_{h \to 0} \frac{1}{\sqrt{x+h} + \sqrt{x}} = \frac{1}{2\sqrt{x}}$$

For the most part, of course, no one really computes derivatives like that. We have theorems like

$$D_x[af(x) + bg(x)] = af'(x) + bg'(x)$$

$$D_x[f(x)g(x)] = f'(x)g(x) + f(x)g'(x) \text{ (multiplication rule)}$$
$$D_x[f(g(x))] = f'(g(x))f'(x) \text{ (chain rule)}$$
$$D_x[f^{-1}(f(x))] = \frac{1}{f'(x)}$$

as well as the derivatives of specific functional forms

$$D_x x^k = ke^{k-1}$$

$$D_x e^x = e^x$$

$$D_x \log(x) = \frac{1}{x}$$

and so on. This allows us to compute many fairly "complicated" derivatives by grinding through the above rules. But a notable feature of economics is that we are fundamentally unsure of what functional forms we should be using, despite the fact that we know a reasonable amount about what they "look" like. These qualitative features are often expressed in terms of derivatives. For example, it is typically assumed that a consumer's benefit from a good is positive, marginal benefit is positive, but marginal benefit is decreasing. In short, $v(q) \geq 0$, $v'(q) \geq 0$ and $v''(q) \leq 0$. A firm's total costs are typically positive, marginal cost is positive, and marginal cost is increasing. In short, $C(q) \geq 0$, $C'(q) \geq 0$, and $C''(q) \geq 0$. By specifying our assumptions this way, we are being precise as well as avoiding the arbitrariness of assuming that a consumer has $a \log(q)$ preferences for some good, but $1 - e^{-bq}$ preferences for another.

3.5.1 Non-differentiability

How do we recognize non-differentiability? Consider

$$f(x) = |x| = \begin{cases} -x & \text{if } x < 0 \\ 0 & \text{if } x = 0 \\ x & \text{if } x > 0 \end{cases},$$

the *absolute value of* x, what is the derivative, $f'(x)$? For $x < 0$, the function is just $f(x) = -x$, which has derivative $f'(x) = -1$. For $x > 0$, the function is just $f(x) = x$, which has derivative $f'(x) = 1$. But what about at zero? First, let's define the *derivative of* f *at* x *from the left* as

$$f'(x^-) = \lim_{h \uparrow 0} \frac{f(x+h) - f(x)}{h}$$

and the *derivative of* f *at* x *from the right* as

$$f'(x^+) = \lim_{h \downarrow 0} \frac{f(x+h) - f(x)}{h}$$

Note that:

Theorem 3.5.1 *A function is differentiable at a point* x *if and only if its one-sided derivatives exist and are equal.*

Then for $f(x) = |x|$ with $x = 0$, we have

$$f'(0^+) = \lim_{h \downarrow 0} \frac{|x+h| - |0|}{h} = \lim_{h \to 0} \frac{h}{h} = 1$$

and

$$f'(0^-) = \lim_{h \uparrow 0} \frac{|0+h| - |0|}{h} = \lim_{h \to 0} \frac{|h|}{h} = -1$$

So we could hypothetically assign any number from -1 to 1 to be the derivative of $|x|$ at zero. In this case, we say that $f(x)$ is *non-differentiable at* x, since the tangent line to the graph of $f(x)$ is not unique — people often say there is a "corner" or "kink" in the graph of $|x|$ at zero. We already computed

$$D_x \sqrt{x} = \frac{1}{2\sqrt{x}}$$

Obviously, we can't evaluate this function for $x < 0$ since $\sqrt{x}$ is only defined for positive numbers. For $x > 0$, the function is also well behaved. But at zero, we have

$$D_x \sqrt{0} = \frac{1}{2 * 0}$$

which is undefined, so the derivative *fails to exist* at zero. So, if you want to show a function is non-differentiable, you need to show that the derivatives from the left and from the right are not equal, or that the derivative fails to exist.

3.6 Taylor Series

It turns out that the sequence of derivatives of a function

$$f'(x), f''(x), f'''(x), ..., f^{(k)}(x), ...$$

generally provides enough information to recover the function, or approximate it "as well as we need" near a particular point using only the first k terms.

Definition 3.6.1 *The k-th order Taylor polynomial of $f(x)$ based at x_0 is*

$$f(x) = \underbrace{f(x_0) + f'(x_0)(x-x_0) + f''(x_0)\frac{(x-x_0)^2}{2} + ... + f^{(k)}(x_0)\frac{(x-x_0)^k}{k!}}_{k\text{-th order approximation of } f} + \underbrace{f^{(k+1)}(c)\frac{(x-x_0)^{k+1}}{(k+1)!}}_{\text{Remainder term}}$$

where c is between x and x_0.

Example The second-order Taylor polynomial of $f(x)$ based at x_0 is

$$f(x) = f(x_0) + f'(x_0)(x-x_0) + \frac{1}{2}f'(c)(x-x_0)^2$$

For $f(x) = e^x$ with $x_0 = 1$, we have

$$f(x) = e + e(x-1) + e\frac{(x-1)^2}{2} + e^c\frac{(x-1)^3}{6}$$

while for $x_0 = 0$ we have

$$f(x) = 1 + x + \frac{x^2}{2} + e^c\frac{x^3}{6}$$

For $f(x) = x^5 + 7x^2 + x$ with $x_0 = 3$, we have

$$f(x) = 309 + 448(x-3) + 554\frac{(x-3)^2}{2} + 60c^2\frac{(x-3)^3}{6}$$

while for $x_0 = 10$ we have

$$f(x) = 100,710 + 50,141(x-10) + 20,014\frac{(x-10)^2}{2} + 60c^2\frac{(x-10)^3}{6}$$

For $f(x) = \log(x)$ with $x_0 = 1$, we have

$$f(x) = (x-1) + \frac{(x-1)^2}{2} + \frac{2}{6c^3}(x-1)^3$$

$\blacksquare$

So while the Taylor series with the remainder/error term is an exact approximation, dropping the approximation introduces error. We often simply work with the approximation and claim that if we are "sufficiently close" to the base point, it won't matter. Or, we will use a Taylor series to expand a function in terms of its derivatives, perform some calculations, and then take a limit so that the error vanishes. Why are these claims valid? Consider the second-order Taylor polynomial,

$$f(x) = f(x_0) + f'(x_0)(x-x_0) + f''(x_0)\frac{(x-x_0)^2}{2} + f'''(c)\frac{(x-x_0)^3}{6}$$

This equality is exact when we include the remainder term, but not when we drop it. Let

$$\hat{f}(x) = f(x_0) + f'(x_0)(x-x_0) + f''(x_0)\frac{(x-x_0)^2}{2}$$

be our *second-order approximation of f around x_0*. Then the *approximation error*

$$\left|f(x) - \hat{f}(x)\right| = \left|f'''(c)\frac{(x-x_0)^3}{6}\right|$$

is just a constant $|f'''(c)|/6$ multiplied by $|(x-x_0)^3|$. Therefore, we can make the error arbitrarily small (less than any $\varepsilon > 0$) by making x very close to x_0:

$$\frac{|f'''(x_0)|}{6}|(x-x_0)^3| < \varepsilon \longrightarrow |x-x_0| < \left(\frac{\varepsilon}{|f'''(x_0)|}\right)^3$$

We write this as

$$f(x) = f(x_0) + f'(x_0)(x-x_0) + f''(x_0)\frac{(x-x_0)^2}{2} + o(h^3)$$

where $h = x - x_0$, or that the error is *order h-cubed*. This is understood to mean that if $h = x - x_0$ is small enough, then $f(x) - f(x_0)$ will be as small as desired. This is important for maximization theory because we will often want to use low-order Taylor polynomials around a local maximum x^*, and we need to know that if x is close enough to x^*, the approximation will satisfy $f(x^*) \geq f(x)$ (why is this important?).

3.7 Partial Derivatives

Most functions of interest to us are not a function of a single variable, but many. As a result, even though we're focused on maximization where the choice variable is one-dimensional, it helps to introduce partial derivatives so we can study how solutions and payoffs vary in terms of variables outside the agent's control.

For example, a firm's profit function

$$\pi(q) = pq - \frac{c}{2}q^2$$

is really a function of q, p and c, or $\pi(q, p, c)$. We will need to differentiate not just with respect to q, but also p and c.

Definition 3.7.1 *Let* $f(x_1, x_2, ..., x_n) : \mathbb{R}^n \to \mathbb{R}$. *The* partial derivative of $f(x)$ with respect to x_i *is*

$$\frac{\partial f(x)}{\partial x_i} = \lim_{h \to 0} \frac{f(x_1, ..., x_i + h, x_n) - f(x_1, ..., x_i, ..., x_n)}{h}$$

The gradient *is the vector of partial derivatives*

$$\nabla f(x) = \begin{pmatrix} \dfrac{\partial f(x)}{\partial x_1} \\ \dfrac{\partial f(x)}{\partial x_2} \\ \vdots \\ \dfrac{\partial f(x)}{\partial x_N} \end{pmatrix}$$

Since this notation can become cumbersome — especially when we differentiate multiple times with respect to different variables x_i and then x_j, and so on — we often write

$$\frac{\partial f(x)}{\partial x_i} = f_{x_i}(x)$$

or

$$\frac{\partial^2 f(x)}{\partial x_j \partial x_i} = f_{x_j x_i}(x)$$

Example Consider a simple profit function

$$\pi(q, p, c) = pq - \frac{c}{2}q^2$$

Then

$$\frac{\partial \pi(q, p, c)}{\partial c} = -\frac{1}{2}q^2$$

and

$$\frac{\partial \pi(q, p, c)}{\partial p} = q$$

and the gradient is

$$\nabla \pi(q, p, c) = \left(\frac{\partial \pi(q, p, c)}{\partial q}, \frac{\partial \pi(q, p, c)}{\partial p}, \frac{\partial \pi(q, p, c)}{\partial c} \right) = \left(q - cq, q, -\frac{1}{2}q^2 \right)$$

■

The partial derivative with respect to x_i holds all the other variables $(x_1, ..., x_{i-1}, x_{i+1}, ..., x_n)$ constant and only varies x_i slightly, exactly like a one-dimensional derivative where the other arguments of the function are treated as constants.

3.7.1 Differentiation with Multiple Arguments and Chain Rules

Recall that the one-dimensional chain rule is that, for any two differentiable functions $f(x)$ and $g(y)$,

$$Dg(f(x)) = g'(f(x))f'(x)$$

Of course, since a partial derivative is just a regular derivative where all other arguments are held constant, it's true that

$$\frac{\partial g(y_1, y_2, ..., f(x_i), ..., y_N)}{\partial x_i} = \frac{\partial g(...)}{\partial y_i} f'(x_i)$$

But we run into problems when we face a function $f(y_1, ..., y_N)$ where many of the variables y_i are functions that depend on some other, common variable, c. For example, consider $f(x(c), c)$. The c term shows up multiple places, so it is not immediately obvious how to differentiate with respect to c.

Let $g(c) = f(x(c), c)$, and consider totally differentiating $g(c)$ with respect to c:

$$g'(c) = \frac{df(x(c), c)}{dc} = \lim_{h \to 0} \frac{f(x(c+h), c+h) - f(x(c), c)}{h}$$

This limit looks incalculable, since both arguments are varying at the same time. Consider using a Taylor series to expand the first term in x at $x(c)$ as

$$f(x(c+h), c+h) = f(x(c), c+h) + \frac{\partial f(\xi, c+h)}{\partial x}(x(c+h) - x(c))$$

where ξ is between $x(c)$ and $x(c+h)$. Let's now expand the first term on the right hand side in c at c, as

$$f(x(c), c+h) = f(x(c), c) + \frac{\partial f(x(c), \zeta)}{\partial c} h$$

where ζ is between c and $c+h$. Inserting the second equation into the first, we get

$$f(x(c+h), c+h) = f(x(c), c) + \frac{\partial f(x(c), \zeta)}{\partial c} h + o(z_2^2) + \frac{\partial f(\xi, c+h)}{\partial x}(x(c+h) - x(c)) + o(z_1^2)$$

Since ξ is between $x(c+h)$ and $x(c)$, it must tend to $x(c)$ as $h \to 0$; since ζ is between c and $c+h$, it must tend to c as $h \to 0$. Then re-arranging and dividing by h yields

$$\frac{f(x(c+h), c+h) - f(x(c), c)}{h} = \frac{\partial f(x(c), \zeta)}{\partial c} + \frac{\partial f(\xi, c+h)}{\partial x} \frac{x(c+h) - x(c)}{h}$$

The left-hand side is almost the derivative of $g(c)$, we just need to take limits with respect to h. Taking the limit as $h \to 0$ then yields

$$g'(c) = \frac{df(x(c), c)}{dc} = f_c(x(c), c) + f_x(x(c), c)\frac{\partial x(c)}{\partial c}$$

So we work argument by argument, partially differentiating all the way through using the chain rule, and then summing all the resulting terms. For example,

$$\frac{d}{dc} g(x_1(c), x_2(c), ..., x_N(c), c) = \left(\sum_{i=1}^{N} g_{y_i}(x(c), c) \frac{\partial x_i(c)}{\partial c} \right) + \frac{\partial g(x(c), c)}{\partial c}$$

Exercises

1. Write out the first few terms of the following sequences, find their limits if the sequence converges, and find the suprema, and infima:

$$x_n = \frac{(-1)^n}{n}$$

$$x_n = \frac{\sqrt{n}}{n}$$

$$x_n = \frac{\sqrt{n+1} - \sqrt{n}}{\sqrt{n+1} + \sqrt{n}}$$

$$x_n = \sin\left(\frac{\pi(n-1)}{2}\right)$$

(Hint: Does this sequence have multiple convergent sub-sequences? Argue that a sequence with multiple convergent sub-sequences that have different limits cannot be convergent.)

$$x_n = \left(\frac{1}{n}\right)^{1/n}$$

(Hint: Show that x_n is an increasing sequence, and then argue that $\sup_n (1/n)^{1/n} = 1$. Then $x_n \to 1$, right?)

2. Give an example of a sequence x_n and a function $f(x)$ so that $x_n \to c$, but $\lim_n f(x_n) \neq f(c)$.

3. Suppose x^* is a local maximizer of $f(x)$. Let $g()$ be a strictly increasing function. Is x^* a local maximizer of $g(f(x))$? Suppose x^* is a local minimizer of $f(x)$. What kind of transformations $g()$ ensure that x^* is also a local minimizer of $g(f(x))$? Suppose x^* is a local maximizer of $f(x)$. Let $g()$ be a strictly decreasing function. Is x^* a local maximizer or minimizer of $g(f(x))$?

4. Rewrite the proof of the extreme value theorem for minimization, rather than maximization.

5. A function is *Lipschitz continuous on* $[a,b]$ if, for all x' and x'' in $[a,b]$, $|f(x') - f(x'')| < K|x' - x''|$, where K is finite. Show that a Lipschitz continuous function is continuous. Provide an example of a continuous function that is not Lipschitz continuous. How is the Lipschitz constant K related to the derivative of $f(x)$?

6. Using Matlab or Excel, numerically compute first- and second-order approximations of the exponential function with $x_0 = 0$ and $x_0 = 1$. Graph the approximations and the approximation error as you move away from each x_0. Do the same for the natural log function with $x_0 = 1$ and $x_0 = 10$. Explain whether the second- or third-order approximation does better, and how the performance degrades as x moves away from x_0 for each approximation.

7. Suppose $f'(x) > g'(x)$. Is $f(x) > g(x)$? Prove that if $f'(x) > g'(x) > 0$, there exists a point x_0 such that for $x > x_0$, $f(x) \geq g(x)$.

8. Explain when the derivatives of $f(g(x))$, $f(x)g(x)$, and $f(x)/g(x)$ are positive for all x or strictly negative for all x.

9. Consider a function $f(c,d) = g(y_1(c,d), y_2(d), c)$. Compute the partial derivatives of $f(c,d)$ with respect to c and d. Repeat with $f(a,b) = g(y_1(z(a), h(a,b)), y_2(w(b)))$.

Proofs

Bolzano-Weierstrass Theorem If a sequence is bounded, all its terms are contained in a set $I = [a, b]$. Take the first term of the sequence, x_1, and let it be x_{k_1}. Now, split the interval $I = [a, b]$ into $[a, 1/2(b-a)]$ and $[1/2(b-a), b]$. One of these subsets must have an infinite number of terms, since x_n is an infinitely long sequence. Pick that subset, and call it I_1. Select any member of the sequence in I_1, and call it x_{k_2}. Now split I_1 in half. Again, one of the subsets of I_1 has an infinite number of terms, so pick it and call it I_2. Proceeding in this way, splitting the sets in half and picking an element from the set with an infinite number of terms in it, we construct a sequence of sets I_k and a subsequence x_{n_k}.

Now note that the length of the sets I_k is

$$(b-a)\left(\frac{1}{2}\right)^k \to 0$$

So the distance between the terms of the sub-sequence x_{n_k} and $x_{n_{k+1}}$ cannot be more than $(b-a)(1/2)^k$. Since this process continues indefinitely and $(b-a)(1/2)^k \to 0$, there will be a limit term $\bar{x}$ of the sequence [1]. Then for some k we have $|x_{n_k} - \bar{x}| < (1/2)^k$, which can be made arbitrarily small by making k large, or for any $\varepsilon > 0$ if $k \geq K$ then $|x_{n_k} - \bar{x}| < \varepsilon$ for $K \geq \log(\varepsilon)/\log(1/2)$. ∎

Sequential Criterion for Continuity Suppose $f(x)$ is continuous at c. Then for all $\varepsilon > 0$, there is a $\delta > 0$ so that $|x - c| < \delta$ implies $|f(x) - f(c)| < \varepsilon$. Take any sequence $x_n c$. This implies that for all $\eta > 0$ there is some $n > N$ so that $|x_n - c| \leq \delta$. Then $|x_n - c| < \delta$, so that $|f(x_n) - f(c)| < \varepsilon$. Therefore, $\lim f(x_n) = f(c)$.

Take any sequence $x_n \to c$. Then for all $\varepsilon > 0$ there exists an N' so that $n \geq N'$ implies $|f(x_n) - f(c)| < \varepsilon$. Since $x_n \to c$, for all $\delta > 0$ there is an N'' so that $n \geq N'$ implies $|x_n - c| < \delta$. Let $N = \max\{N', N''\}$. Then for all $\varepsilon > 0$, there is a $\delta > 0$ so that $|x_n - c| < \delta$ implies $|f(x_n) - f(c)| < \varepsilon$. Therefore $f(x)$ is continuous. ∎

Questions about sequences and the extreme value theorem:

7. Prove that a convergent sequence is bounded. (You will need the inequality $|y| < c$ implies $-c \leq y \leq c$. Try to bound the "tail" terms as $|x_n| < |\varepsilon + x_N|$, $m > N$, and then argue the sequence is bounded by the constant $M = \max\{x_1, x_2, ..., x_N, |\varepsilon + x_N|\}$.)

8. Prove that a convergent sequence has a unique limit. (Start by supposing that x_n converges to x' and x''. Then use the facts that $|y| \leq c$ implies $-c \leq y \leq c$ and $|a+b| \leq |a| + |b|$ to show that $|x'' - x'| < |\varepsilon/2| + |x_n - x''|$, and $|x_n - x''|$ can be made less $\varepsilon/2$.)

9. The extreme value theorem says that if $f : [a, b] \to \mathbb{R}$ is continuous, then a maximizer of $f(x)$ exists in $[a, b]$. Is it necessary for $f(x)$ to be continuous for this result to be true (provide an example to explain your answer). A function $f : [a, b] \to \mathbb{R}$ is *upper semi-continuous* if, for all $\varepsilon > 0$, there exists a $\delta > 0$ so that if $|x - c| < \delta$, then

$$f(x) \leq f(c) + \varepsilon$$

(a) Sketch some upper semi-continuous functions. (b) Show that upper semi-continuity is equivalent to: For any sequence $x_n \to c$,

$$\lim_{n\to\infty} \sup_{k \geq n} f(x) \leq f(c)$$

[1] This part of the proof is a little loose. The existence of $\bar{x}$ is ensured by the Cauchy criterion for convergence, which is covered in exercise 9

(c) Use the sequential criterion for upper semi-continuity in part b to show that the extreme value theorem extends to upper semi-continuous functions.

10. A sequence is *Cauchy* if, for all $\varepsilon > 0$ and all $n, m \geq N$, $|x_n - x_m| < \varepsilon$. Show that a sequence $x_n \to \bar{x}$ if and only if it is Cauchy. The bullet points below sketch the proof for you, provide the missing steps:

- To prove a convergent sequence is Cauchy: (Easy part)
  - Add and subtract $\bar{x}$ inside $|x_n - x_m|$, then use the triangle inequality, $|a + b| < |a| + |b|$. Lastly, use the definition of a convergent sequence.
- To prove a Cauchy sequence converges: (Hard part)
  - Show that a Cauchy sequence is bounded. (You will need the inequality $|y| < c$ implies $-c \leq y \leq c$. Try to bound the "tail" terms as $|x_m| < |\varepsilon + x_N|$, $m > N$, and then argue the sequence is bounded by the constant $M = \max\{x_1, x_2, ..., x_N, |\varepsilon + x_N|\}$.)
  - Use the Balzano-Weierstrass theorem to argue that a Cauchy sequence then has a convergent subsequence with limit $\bar{x}$.
  - Show that $x_n \to \bar{x}$, so that the sequence converges to $\bar{x}$. Suppose there is a sub-sequence x_{n_k} of x_n that does not converge to $\bar{x}$, and show that this leads to a contradiction (again, the inequality $|y| \leq c$ implies $-c \leq y \leq c$ is useful. Compare the non-convergent subsequence with the convergent one, along with the definition of a Cauchy sequence for points from the two sequences.).

Cauchy sequences can be very useful when you want to prove a sequence converges, but have no idea what the exact limit is. One example is studying sequences of decision variables in macroeconomics, where, for example, a choice of capital decisions k_t is generated by the economy each period t, but it is unclear whether this sequences converges or diverges.

Chapter 4

Necessary and Sufficient Conditions for Maximization

While existence results are useful in terms of helping us understand which maximization problems have answers at all, they do little to help us *find* maximizers. Since analyzing any economic model relies on finding the optimal behavior for each agent, we need a method of finding maximizers and determining when they are unique.

4.1 First-Order Necessary Conditions

We start by asking the question, "What criteria must a maximizer satisfy?" This throws away many possibilities, and narrows attention on a candidate list. This does not mean that every candidate is indeed a maximizer, but merely that if a particular point is *not* on the list, then it *cannot* be a maximizer. These criteria are called *first-order necessary conditions*.

Example Suppose a price-taking firm has total costs $C(q)$, and gets a price p for its product. Then its profit function is

$$\pi(q) = pq - C(q)$$

How can we find a maximizer of $\pi()$? Let's start by differentiating with respect to q:

$$\pi'(q) = p - C'(q)$$

Since $\pi'(q)$ measures the rate of change of q, we can increase profits if $p - C'(q) > 0$ at q by increasing q a little bit. On the other hand, if $p - C'(q) < 0$, we can raise profits by decreasing q a little bit. The only place where the firm has no incentive to change its decision is where $p = C'(q^*)$.
■

Note that the logic of the above argument is that *if* q^* is a maximizer of $\pi(q)$, *then* $\pi'(q^*) = 0$. We are using a property of a maximizer to derive conditions it must satisfy, or necessary conditions. We can make the same argument for any maximization problem:

Theorem 4.1.1 *If x^* is a local maximum of $f(x)$ and $f(x)$ is differentiable at x^*, then $f'(x^*) = 0$.*

Proof Suppose x^* is a local maximum of $f()$. If $f'(x^*) > 0$, then we could take a small step to $x^* + h$, and the Taylor series would be

$$f(x^* + h) = f(x^*) + f'(x^*)h + o(h^2)$$

implying that for very small h,

$$f(x^* + h) - f(x^*) = f'(x^*)h > 0$$

so that $x^* + h$ gives a higher value of $f()$ than x^*, which would be a contradiction. So $f'(x^*)$ cannot be strictly greater than zero.

Suppose x^* is a local maximum of $f()$. If $f'(x^*) < 0$, then we could take a small step to $x^* - h$, and the Taylor series would be

$$f(x^* - h) = f(x^*) - f'(x^*)h + o(h^2)$$

implying that for very small h,

$$f(x^* - h) - f(x^*) = -f'(x^*) > 0$$

so that $x^* - h$ gives a higher value of $f()$ than x^*, which would be a contradiction. So $f'(x^*)$ cannot be strictly less than zero.

That leaves $f'(x^*) = 0$ as the only time when we can't improve $f()$ by taking a small step away from x^*, so that x^* must be a local maximum. $\blacksquare$

So the basic candidate list for an (unconstrained, single-dimensional maximization problem) boils down to

- The set of points at which $f(x)$ is not differentiable.
- The set of *critical points*, where $f'(x) = 0$.

Example A consumer with utility function $u(q, m) = b\log(q) + m$ over the market good q and money spent on other goods m faces budget constraint $w = pq + m$. The consumer wants to maximize utility.

We can rewrite the constraint as $m = w - pq$, and substitute it into $u(q, m)$ to get $b\log(q) + w - pq$. Treating this as a one-dimensional maximization problem,

$$\max_q b\log(q) + w - pq$$

our FONCs are

$$\frac{b}{q^*} - p = 0$$

or

$$q^* = \frac{b}{p}$$

$\blacksquare$

Note however, that the set of critical points potentially includes some local minimizers. Why? If x^* is a maximizer of $-f(x)$, then $-f'(x^*) = 0$; but then x^* is a minimizer of $f(x)$, and $f'(x*) = 0$. So being a critical point is not enough to guarantee that a point is a maximizer.

Example Consider the function

$$f(x) = -\frac{1}{4}x^4 + \frac{c^2}{2}x^2$$

The FONC is

$$f'(x) = -x^3 + c^2 x = 0$$

One critical point is $x^* = 0$. Dividing by $x \neq 0$ and solving by radicals gives two more: $x^* = \pm c$. So our candidate list has three entries. To figure out which is best, we substitute them back into the objective:

$$f(0) = 0, f(+c) = f(-c) = \frac{1}{4}c^4 > 0$$

So $+c$ and $-c$ are both global maxima (since $f(x)$ is decreasing for $x > c$ and $x < c$, we can ignore all those points). The point $x = 0$ is a local minimum (but not global minimum). $\blacksquare$

On the other hand, we can find critical points that are *neither* maxima nor minima.

Example Let $f(x) = x^3$ on $[-1, 1]$ (a solution to this problem exists, right?). Then

$$f'(x) = 3x^2$$

has only one solution, $x = 0$. So the only critical point of x^3 is zero. But it is neither a maximum nor a minimum on $[-1, 1]$ since $f(0) = 0$, but $f(1) = 1$ and $f(-1) = -1$. This is an example of an *inflection point* or a *saddle point*. $\blacksquare$

However, if we have built the candidate list correctly, then one or more of the points on the candidate list *must* be the global maximizer. If worse comes to worst, we can always evaluate the function $f(x)$ for every candidate on the list, and compare their values to see which does best. We would then have found the global maximizer for sure. (But this might be prohibitively costly, which is why we use second-order sufficient conditions, which are the next topic).

So even though FONCs are useful for building a candidate list (non-differentiable points and critical points), they don't discriminate between maxima, minima, and inflection points. However, we can develop ways of testing critical points to see how they behave locally, called second-order sufficient conditions.

4.2 Second-Order Sufficient Conditions

The idea of *second-order sufficient conditions* is to provide criteria that ensure a critical point is a maximum or minimum. This gives us both a test to see if a point on the candidate list is a local maximum or local minimum, as well as provides more information about the behavior of a function near a local maximum in terms of calculus.

Theorem 4.2.1 *If $f'(x^*) = 0$ and $f''(x^*) < 0$, then x^* is a local maximum of $f()$.*

Proof The Taylor series at $f(x^*)$ is

$$f(x) = f(x^*) + f'(x^*)(x - x^*) + f''(x^*)\frac{(x - x^*)^2}{2} + o(h^3)$$

Since $f'(x^*) = 0$, we have

$$f(x) = f(x^*) + f''(x^*)\frac{(x - x^*)^2}{2} + o(h^3)$$

and re-arranging yields

$$f(x^*) - f(x) = -f''(x^*)\frac{(x - x^*)^2}{2} - o(h^3)$$

so for $h = x - x^*$ very close to zero, we get

$$f(x^*) - f(x) = -f''(x^*)\frac{(x - x^*)^2}{2} > 0$$

and

$$f(x^*) > f(x)$$

so that x^* is a local maximum of $f(x)$. ∎

This is the standard proof of the SOSCs, but it doesn't give much geometric intuition about what is going on. Using one-sided derivatives, the second derivative can always be written as

$$f''(x) = \lim_{h\to 0} \frac{f'(x+h) - f'(x)}{h} = \lim_{h\to 0} \frac{\dfrac{f(x+h)-f(x)}{h} - \dfrac{f(x)-f(x-h)}{h}}{h}$$

which equals

$$f''(x) = \lim_{h\to 0} \frac{f(x+h) + f(x-h)) - 2f(x)}{h^2}$$

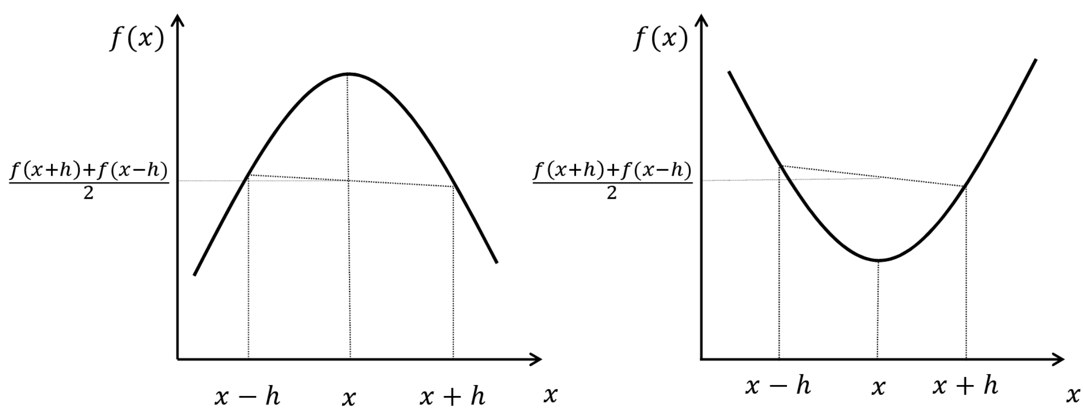
or

$$f''(x) = \lim_{h\to 0} 2\frac{\dfrac{f(x+h)+f(x-h)}{2} - f(x)}{h^2}$$

Now, the term

$$\frac{f(x+h)+f(x-h)}{2}$$

is the average of the points one step above and below $f(x)$. So if $(f(x+h)+f(x-h))/2 < f(x)$, the average of the function values at $x+h$ and $x-h$ is less than the value at x, so the function must locally look like a "hill". If $(f(x+h)+f(x-h))/2 > f(x)$, then the average of the function values at $x+h$ and $x-h$ is above the function value at x, and the function must locally look like a "valley". This is the real intuition for the second-order sufficient conditions: The second derivative is testing the "curvature" of the function to see whether it's a hill or a valley.



Second-Order Sufficient Conditions

If $f''(x) = 0$, however, we can't conclude anything about a critical point. Recall that if $f(x) = x^3$, then $f'(x) = 3x^2$ and $f''(x) = 6x$. Evaluated at zero, the critical point $x = 0$ gives $f''(0) = 0$. So an indeterminate second derivative provides no information about whether we are at a maximum, minimum or inflection point.

Example Consider a price-taking firm with cost function

$$\max_q pq - C(q)$$

To maximize profit, its FONCs are

$$p - C'(q^*) = 0$$

and its SOSCs are

$$-C''(q^*) < 0$$

So as long as $C''(q^*) > 0$, a critical point is a maximum. For example $C(q) = \dfrac{c}{2}q^2$ satisfies $C''(q) = c$, so the SOSCs are always satisfied $\blacksquare$

Example Consider a consumer with utility function $u(q, m) = v(q) + m$ and budget constraint $w = pq + m$. The consumer then maximizes

$$\max_q v(q) + w - pq$$

yielding FONCs

$$v'(q^*) - p = 0$$

and SOSCs

$$v''(q^*) < 0$$

So as long as $v''(q^*) < 0$ a critical point q^* is a local maximum. For example, $v(q) = b\log(q)$ satisfies this condition. $\blacksquare$

Example Recall the function

$$f(x) = -\frac{1}{4}x^4 + \frac{c^2}{2}x^2$$

The FONC is

$$f'(x) = -x^3 + c^2 x = 0$$

And it had three critical points, 0 and $\pm c$. The second derivative is

$$f''(x^*) = -3x^2 + 2c^2 < 0$$

For $x^* = 0$, we get $f''(0) = 2c^2 > 0$, so it is a local minimum, not maximum. For $x^* = \pm c$, we get $f''(\pm c) = -3c^2 + 2c^2 = -c^2 < 0$, so these points are both local maxima. $\blacksquare$

Example Suppose an agent consumes in period 1 and period 2, with utility function

$$u(c_1, c_2) = \log(c_1) + \delta \log(c_2)$$

where $c_1 + s = y_1$ and $Rs + y_2 = c_2$, where $R > 1$ is the interest rate plus 1. Substituting the constraints into the objective yields the problem

$$\max_s \log(y_1 - s) + \delta \log(Rs + y_2)$$

Maximizing over s yields the FONC

$$-\frac{1}{y_1 - s^*} + \delta R \frac{1}{Rs^{*2} + y_2} = 0$$

and the SOSC is

$$-\frac{1}{(y_1 - s^*)^2} + \delta R^2 \frac{-1}{(Rs^{*2} + y_2)^2} < 0$$

which is automatically satisfied, since for any s — not just a critical point — we have

$$-\frac{1}{(y_1 - s)^2} + \delta R^2 \frac{-1}{(Rs^2 + y_2)^2} < 0$$

$\blacksquare$

It's a "nice" feature that the SOSCs are automatically satisfied in the previous example. If we could find general features of functions that guarantee this, it would make our lives much easier. In particular, it turned out that $f''(x) < 0$ for all x, not just at a critical point x^*. This is the special characteristic of a *concave function*.

We can say a bit more about local maximizers using a similar approach.

Theorem 4.2.2 *If x^* is a local maximum of $f()$ and $f'(x^*) = 0$, then $f''(x^*) \leq 0$.*

Proof If x^* is a local maximum, then the Taylor series around x^* is

$$f(x) = f(x^*) + f'(x^*)(x - x^*) + f''(x^*)\frac{(x - x^*)^2}{2} + o(h^3)$$

Using the FONCs and re-arranging yields

$$f(x^*) - f(x) = -f''(x^*)\frac{(x - x^*)^2}{2} + o(h^3)$$

Since x^* is a local maximum, $f(x^*) \geq f(x)$, and this implies that for x close enough to x^*,

$$f''(x^*) \leq 0$$

∎

These are called *second-order necessary conditions*, since they follow by necessity from the fact that x^* is a local maximum and critical point (can you have a local maximum that is not a critical point?). Exercise 6 asks you to explain the difference between Theorem 4.2.1 and Theorem 4.2.2. This is one of these subtle points that will bother you in the future if you don't figure out now.

Exercises

1. Derive FONCs and SOSCs for the firm's profit-maximization problem for the cost functions $C(q) = \frac{c}{2}q^2$ and $C(q) = e^{cx}$.

2. When an objective function $f(x, y)$ depends on two controls, x and y, and is subject to a linear constraint $c = ax + by$, the problem can be simplified to a one-dimensional program by solving the constraint in terms of one of the controls,

$$y = \frac{c - ax}{b}$$

and substituting it into the objective to get

$$\max_x f\left(x, \frac{c - ax}{b}\right)$$

Derive FONCs and SOSCs for this problem. What assumptions guarantee that the SOSC's hold?

3. Derive FONCs and SOSCs for the consumer's utility-maximization problem for the benefit functions $v(q) = b\log(q)$, $v(q) = 1 - e^{-bq}$, and $v(q) = bq - \frac{c}{2}q^2$.

4. For a monopolist with profit function

$$\max_q p(q)q - C(q)$$

where $C(q)$ is an increasing, convex function, derive the FONCs and SOSCs. Provide the best sufficient condition you can think of for the SOSCs to hold for any increasing, convex cost function $C(q)$.

5. Explain the difference between Theorem 4.2.1 and Theorem 4.2.2 both (i) as completely as possible, and (ii) as briefly as possible. Examples can be helpful.

6. Prove that for a function $f : \mathbb{R}^n \to \mathbb{R}$, a point x^* is a local maximum only if $\nabla f(x) = 0$. Suppose a firm hires both capital K and labor L to produce output through technology $F(K, L)$, where p is the price of its good, w is the price of hiring a unit of labor, and r is the price of hiring a unit of capital. Derive FONCs for the firm's profit maximization problem.

7. For a function $f : \mathbb{R}^n \to \mathbb{R}$ and a critical point x^*, is it sufficient that $f(x)$ have a negative second partial derivative for each argument x_i

$$\frac{\partial^2 f(x^*)}{\partial x_i^2} < 0$$

for x^* to be a local maximum of $f(x)$? Prove or disprove with a counter-example. (Try using Matlab or Excel to plot the family of quadratic forms

$$f(x_1, x_2) = a_1 x_1 + a_2 x_2 + \frac{b_1}{2} x_1^2 + \frac{b_2}{2} x_2^2 + b_3 x_1 x_2$$

and experiment with different coefficients to see if the critical point is a maximum or not.)

8. Suppose you have a hard maximization problem that you cannot solve by hand, or even in a straightforward way on a computer. How might you proceed? (i) Write out a second-order Taylor series around an *initial guess*, x_0, to the point x_1. Now derive an expression for $f(x_1) - f(x_0)$, and maximize the difference (you are using a local approximation of the function to maximize the increase in the function's value as you move from x_0 to x_1). Solve the first-order necessary condition in terms of x_1. (ii) If you replace x_0 with x_k and x_1 with x_{k+1}, repeating this procedure will generate a sequence of guesses x_k. When do you think it converges to the true x^*? This numerical approach to finding maximizers is called *Newton's method*.

Chapter 5

The Implicit Function Theorem

The quintessential example of an economic model is a partial equilibrium market with a demand equation, expressing how much consumers are willing to pay for an amount q as a function of weather w,

$$p = f^d(q, w)$$

and a supply equation, expressing how much firms require to produce an amount q as a function of technology t,

$$p = f^s(q, t)$$

At a market-clearing equilibrium (p^*, q^*), we have

$$f^d(q^*, w) = p^* = f^s(q^*, t)$$

But now we might ask, how do q^* and p^* change when w and t change? If the firms' technology t improves, how is the market affected? If the weather w improves and more consumers want to travel, how does demand shift? Since we rarely know much about $f^d()$ and $f^s()$, we want to avoid adding additional, unneeded assumptions to the framework. The implicit function theorem allows us to do this.

In economics, we make a distinction between *endogenous variables* — p^* and q^* — and *exogenous variables* — w and t. An exogenous variable is something like the weather, where no economic agent has any control over it. Another example is the price of lettuce if you are a consumer: Refusing to buy lettuce today probably has no impact on the price, so it essentially beyond your control, and can be treated as a constant. An *endogenous variable* is one that is determined "within the system". For example, the weather is out of your control, but you can choose whether to go to the beach or go to the movies. Prices might be beyond a consumer's control, but the consumer can still choose what bundle of goods to purchase.

A "good model" is one which judiciously picks what behavior to explain — endogenous variables — in terms of economic circumstances that are plausibly beyond the agents' control — exogenous variables, in the clearest and simplest way possible.

Once a good model is provided, we want to ask, "how do endogenous variables change when exogenous variables change?" The Implicit Function Theorem is our tool for doing this. You might want to see how a change in a tax affects behavior, or how an increase in the interest rate affects capital accumulation. Since data are not available for these "hypothetical" worlds, we need models to explain the relationships between exogenous and endogenous variables, so we can then adjust the exogenous variables in our theoretical laboratory.

Example Consider a partial equilibrium market where consumers have utility function $u(q, m) = v(q) + m$ where $v(q)$ is increasing and concave, and a budget constraint $w = pq + m$. Then consumers maximize

$$\max_q v(q) + w - pq$$

yielding a FONC

$$v'(q^*) - p = 0$$

and an *inverse demand curve* is given by $p = v'(q^D)$. Suppose that firms have increasing, convex costs $C(q)$, and face a tax t for every unit they sell. Then their profit function is

$$\pi(q) = (p - t)q - C(q)$$

The FONC is

$$p - t - C'(q^*) = 0$$

and the *inverse supply curve* is given by $p = t + C'(q^S)$.

A *market-clearing equilibrium* (p^*, q^*) occurs where

$$v'(q^*) = p^* = t + C'(q^*)$$

If we re-arrange this equation, we get

$$v'(q^*) - t - C'(q^*) = 0$$

Now, if we think of the market-clearing quantity $q^*(t)$ as a function of taxes t, we can totally differentiate with respect to t to get

$$v''(q^*(t))\frac{\partial q^*}{\partial t} - 1 - C''(q^*(t))\frac{\partial q^*}{\partial t} = 0$$

and re-arranging yields

$$\frac{\partial q^*}{\partial t} = \frac{1}{v''(q^*(t)) - C''(q^*(t))} < 0$$

So that if t increases, the market-clearing quantity falls (what happens to the market-clearing price?). ∎

The general idea of the above two exercises is called the Implicit Function Theorem.

5.1 The Implicit Function Theorem

Suppose we have an equation

$$f(x, c) = 0$$

Then an *implicit solution* is a function $x(c)$, so that

$$f(x(c), c) = 0$$

We say the variables c are *exogenous variables*, and that the x are *endogenous variables.*

Theorem 5.1.1 (Implicit Function Theorem) *Suppose* $f(x_0, c_0) = 0$ *and* $\frac{\partial f(x_0, c_0)}{\partial x} \neq 0$. *Then there exists a continuous implicit solution* $x(c)$, *with derivative*

$$\frac{\partial x(c)}{\partial c} = -\frac{f_c(x(c), c)}{f_x(x(c), c)}$$

for c *close to* c_0.

Proof If we differentiate $f(x(c), c)$ with respect to c, we get

$$\frac{\partial f(x(c),c)}{\partial x}\frac{\partial x(c)}{\partial c} + \frac{\partial f(x(c),c)}{\partial c} = 0$$

Re-arranging the equation yields

$$\frac{\partial x(c)}{\partial c} = -\frac{\partial f(x(c),c)/\partial c}{\partial f(x(c),c)/\partial x}$$

which exists for all c only if $\partial f(x(c), c)/\partial x \neq 0$ for all c. Since we have computed $\dfrac{\partial x(c)}{\partial c}$, it follows that $x(c)$ is continuous, since all differentiable functions are continuous. ∎

The key step in proving the implicit function theorem is the equation

$$\frac{\partial f(x(c),c)}{\partial x}\frac{\partial x(c)}{\partial c} + \frac{\partial f(x(c),c)}{\partial c} = 0$$

The first term is the *equilibrium effect* or *indirect effect*: The system itself adjusts the endogenous variables to restore equilibrium in the equation $f(x(c), c) = 0$ (for us, this will occur because a change in c causes agents to re-optimize, cancelling out part of the change). The second term is the *direct effect*: By changing the parameters, the nature of the system has changed slightly, so that equilibrium will be achieved in a slightly different manner.

Example Consider a monopolist's profit-maximization problem:

$$\max_q p(q)q - \frac{c}{2}q^2$$

The first-order necessary condition is

$$p'(q)q + p(q) - cq = 0$$

This equation fits our $f(x(c), c)$, with $q(c)$ and c as the endogenous and exogenous parameters. Totally differentiating as in the proof yields

$$p''(q)q\frac{dq}{dc} + 2p'(q)\frac{dq}{dc} - q(c) - c\frac{dq}{dc} = 0$$

and re-arranging a bit yields

$$\frac{dq}{dc}\left(p''(q)q + 2p'(q) - c\right) - q(c) = 0$$

The second term is the direct effect: a small increase in c reduces the monopolist's profits by $-q(c)$. The first term is the equilibrium effect: The monopolist adjusts $q(c)$ to maintain the FONC, and a small change in q changes the FONC by exactly the term in parentheses. But it is unclear how to sign this in general, since the term in parentheses may not immediately mean anything to you (is it positive? negative?). ∎

5.2 Maximization and Comparative Statics

Suppose we have an optimization problem

$$\max_x f(x, c)$$

where c is some parameter the decision-maker takes as given, like the temperature or a price that can't be influenced by gaming the market. Let $x^*(c)$ be a maximizer of $f(x,c)$. Then the FONCs imply

$$\frac{\partial f(x^*(c),c)}{\partial x} = f_x(x^*(c),c) = 0$$

and

$$\frac{\partial^2 f(x^*(c),c)}{\partial x^2} = f_{xx}(x^*(c),c) \leq 0$$

Note that we *know* that $f_x(x^*(c),c) = 0$ and $f_{xx}(x^*(c),c) \leq 0$ since we are *assuming* that $x^*(c)$ is a maximizer.

The FONC looks exactly like the kinds of equations studied in the proof of the implicit function theorem, $f_x(x,c) = 0$. The only difference is that it is generated by a maximization problem, not an abstractly given equation. If we differentiate the FONC with respect to c, we get

$$f_{xx}(x^*(c),c)\frac{\partial x^*(c)}{\partial c} + f_{cx}(x^*(c),c) = 0$$

and solving for $\partial x^*/\partial c$ yields

$$\frac{\partial x^*(c)}{\partial c} = \frac{f_{cx}(x^*(c),c)}{-f_{xx}(x^*(c),c)}$$

This is called the *comparative static of x^* with respect to c*: We are measuring how $x^*(c)$ — the agent's behavior — responds to a change in c — some exogenous parameter outside their control. The SOSC implies that $f_{xx}(x^*(c),c) < 0$, so we know that

$$\text{sign}\left(\frac{\partial x^*(c)}{\partial c}\right) = \text{sign}\ \left(f_{cx}(x^*(c),c)\right)$$

So that $x^*(c)$ is increasing in c if $f_{cx}(x^*(c),c) \geq 0$.

Theorem 5.2.1 *Suppose $x^*(c)$ is a local maximum of $f(x,c)$. Then*

$$\textit{sign}\ \frac{\partial x^*(c)}{\partial c} = \textit{sign}\ f_{cx}(x^*(c),c)$$

Example Recall the monopolist, facing the problem

$$\max_q p(q)q - \frac{c}{2}q^2$$

His FONCs are

$$p'(q^*)q^* + p(q^*) - cq^* = 0$$

and his SOSCs are

$$p''(q^*)q^* + 2p'(q^*) - c < 0$$

We apply the IFT to the FONCS: Treat q^* as an implicit function of c, and totally differentiate to get

$$p''(q^*)q^*\frac{\partial q^*}{\partial c} + 2p'(q^*)\frac{\partial q^*}{\partial c} - c\frac{\partial q^*}{\partial c} - q^* = 0$$

Re-arranging, we get,

$$\underbrace{\left(p''(q^*)q^* + 2p'(q^*) - c\right)}_{\text{SOSCs, <0}}\frac{\partial q^*}{\partial c} - q^* = 0$$

or

$$\frac{\partial q^*}{\partial c} = \frac{q^*}{p''(q^*)q^* + 2p'(q^*) - c} < 0$$

So we use the information from the SOSCs to sign the denominator, giving us an unambiguous comparative static. ∎

Example Suppose an agent consumes in period 1 and period 2, with utility function

$$u(c_1, c_2) = \log(c_1) + \delta \log(c_2)$$

where $c_1 + s = y_1$ and $Rs + y_2 = c_2$, where $R > 1$ is the interest rate plus 1. Substituting the constraints into the objective yields the problem

$$\max_s \log(y_1 - s) + \delta \log(Rs + y_2)$$

Maximizing over s yields the FONC

$$-\frac{1}{y_1 - s^*} + \delta R \frac{1}{Rs^{*2} + y_2} = 0$$

and the SOSC is automatically satisfied, since for any s — not just a critical point — we have

$$-\frac{1}{(y_1 - s)^2} + \delta R^2 \frac{-1}{(Rs^2 + y_2)^2} < 0$$

How does s^* vary with R? Well, I don't want to write it all out again. Let the FONC be the function defined as

$$f(s^*(R), R) = 0$$

Then we know that

$$\frac{\partial s^*}{\partial R} = \frac{f_R(s^*(R), R)}{-f_S(s^*(R), R)}$$

From the SOSCs, $f_S(s^*(R), R)$ is negative, so the denominator is positive, and since

$$f_R(s^*(R), R) = \delta \frac{1}{Rs^* + y_2} - \delta R s^{*2} \frac{1}{(Rs^{*2} + y_2)}$$

We only need to sign $f_R(s^*(R), R)$ to get sign of the expression. It is positive if

$$\delta(Rs^{*2} + y_2) - \delta Rs^{*2} = \delta y_2 > 0$$

which is true. So s^* is increasing in R (and look, we didn't even solve for $s^*(R)$). $\blacksquare$

Example Consider a firm with maximization problem

$$\pi(q) = pq - C(q, t)$$

where t represents the firm's technology. In particular, increasing t reduces the firm's marginal costs for all q, or

$$\frac{\partial}{\partial t} \frac{\partial C(q, t)}{\partial q} = C_{qt}(q, t) < 0$$

The firm's FONC is

$$p - C_q(q^*, t) = 0$$

and the SOSC is

$$-C_{qq}(q^*, t) < 0$$

We can study how the optimal quantity q^* varies with either p or t. Let's start with p. Applying the implicit function theorem, we get

$$\frac{\partial q^*(p)}{\partial p} = \frac{1}{C_{qq}(q^*(p), t)} > 0$$

so the supply curve is upward sloping. If t increases instead, we get

$$\frac{\partial q^*(t)}{\partial t} = \frac{-C_{qt}(q^*,t)}{C_{qq}(q^*(p),t)} > 0$$

so that if technology improves, the supply curve shifts up.

We might ask, what is the cross-partial with respect to both t and p? There's nothing stopping us from pushing this approach further. Differentiating the FONC with respect to p, we get

$$1 - C_{qq}(q^*(t,p),t)\frac{\partial q^*(t,p)}{\partial p} = 0$$

and again with respect to t we get

$$-C_{qqq}(q^*(t,p),t)\frac{\partial q^*(t,p)}{\partial t}\frac{\partial q^*(t,p)}{\partial p} - C_{tqq}(q^*(t,p),t)\frac{\partial q^*(t,p)}{\partial p} - C_{qq}(q^*(t,p),t)\frac{\partial^2 q^*(t,p)}{\partial t \partial p} = 0$$

But now we're in trouble. What's the sign of $C_{qqq}(q^*(t,p),t)$? Can we make any sense of this? This is the kind of game you often end up playing as a theorist. We have a complicated, apparently ambiguous comparative static that we would like to make sense of. The goal now is to figure out what kinds of worlds have unambiguous answers. Let's start by re-arranging to get the comparative static of interest alone and seeing what we can sign with our existing assumptions:

$$\frac{\partial^2 q^*(t,p)}{\partial t \partial p} = \frac{C_{qqq}(q^*(t,p),t)\underbrace{\frac{\partial q^*(t,p)}{\partial t}\frac{\partial q^*(t,p)}{\partial p}}_{+} + C_{tqq}(q^*(t,p),t)\underbrace{\frac{\partial q^*(t,p)}{\partial p}}_{+}}{\underbrace{-C_{qq}(q^*(t,p),t)}_{-}}$$

So if $C_{qqq}(q,t)$ and $C_{tqq}(q,t)$ have the same sign, the cross-partial of q with respect to p and t will be unambiguous. Otherwise, you need to make more assumptions about how variables relate, adopt specific functional forms, or rely on the empirical literature to argue that some quantities are more important than others. ∎

Example As an application of the implicit function theorem, consider the equation

$$f^{-1}(f(x)) = x$$

If we differentiate this with respect to x, we get

$$\frac{df^{-1}(y)}{dy}f'(x) = 1$$

or

$$\frac{df^{-1}(y)}{dy} = \frac{1}{f'(x)}$$

Theorem 5.2.2 (Inverse Function Theorem) *If $f'(x) > 0$ for all x, then $\left(f^{-1}\right)'(y) > 0$ for all y, and if $f'(x) < 0$ for all x, then $\left(f^{-1}\right)'(y) < 0$ for all y. If $f(x) = y$ and $f^{-1}(y) = x$, then*

$$\frac{df^{-1}(y)}{dy} = \frac{1}{f'(x)}$$

This allows us to sign the inverse of a function just from information about the derivative of the function itself. This is actually pretty useful. This exact theorem is a key step in deriving the Nash equilibrium of many pay-as-bid auctions, for example. ∎

Example From the implicit function theorem, we have

$$\frac{\partial x(c)}{\partial c} = -\frac{f_{xc}(x(c),c)}{f_{xx}(x(c),c)}$$

Multiplying by c and dividing by $x(c)$ gives

$$\frac{\partial x(c)}{\partial c}\frac{c}{x(c)} = -\frac{f_{xc}(x(c),c)}{f_{xx}(x(c),c)}\frac{c}{x(c)}$$

which is often written as

$$\frac{\partial \log(x(c))}{\partial \log(c)} = \frac{\%\Delta x(c)}{\%\Delta c} = -\frac{f_{xc}(x(c),c)}{f_{xx}(x(c),c)}\frac{c}{x(c)}$$

This quantity

$$\frac{\partial x(c)}{\partial c}\frac{c}{x(c)} = \frac{\partial \log(x(c))}{\partial \log(c)}$$

is called an *elasticity*. Why are these useful?

Suppose we are comparing the effect of a tax on the gallons of water and bushels of apples traded. We might naively compute the derivatives, but what do the numbers even mean? One good is denominated in gallons, and the other in bushels. We might convert gallons to bushels by comparing the weight of water and apples and coming up with a gallons-to-bushels conversion scale, like fahrenheit to celsius. But again, this is somewhat irrelevant. People consume a large amount of water everyday, while they probably only eat a few dozen apples a year (if that). So we might gather data on usage and improve our conversion scale to have economic significance, rather than just physical significance. This approach, though, seems misguided.

The derivative of the quantity $q(t)$ of water traded with respect to t is

$$\lim_{t' \to t} \frac{q(t') - q(t)}{t' - t}$$

so the numerator is the change in the quantity, measured in gallons, while the denominator is the change in the tax, measured in dollars. If we multiply by $t/q(t)$, we get

$$\frac{q(t') - q(t)}{t' - t}\frac{\text{gallons}}{\text{dollars}} \times \frac{t}{q(t)}\frac{\text{dollars}}{\text{gallons}} = \frac{q(t') - q(t)}{t' - t}\frac{t}{q(t)}$$

so the units "cancel" leaving a *dimensionless quantity*, the elasticity. This can be freely compared across goods since we don't have to keep track of currency, quantity denominations, economic significance of units, and so on. ∎

Exercises

1. A monopolist faces demand curve $p(q,a)$, where a is advertising expenditure, and has costs $C(q)$. Solve for the monopolist's optimal quantity, $q^*(a)$, and explain how the optimal quantity varies with advertising if $p_a(q,a) \geq 0$.

2. Suppose there is a perfectly competitive market, where consumers maximize utility $v(q,w) + m$ subject to budget constraints $w = (p+t)q + m$, where t is a tax on consumption of q, and firms have costs $C(q) = cq$. (i) Characterize a perfectly competitive equilibrium in the market, and show how tax revenue — $tq^*(t,w)$ — varies with t and w. (ii) Suppose w changes — there is a shock to consumer preferences — and the government wants to adjust t to hold tax revenue constant. Use the IFT to show how to achieve this.

3. An agent is trying to decide how much to invest in a safe asset that yields return 1 and a risky asset that returns H with probability p and L with probability $1-p$, where $H > 1 > L$. His budget constraint is $w = pa + S$, where p is the price of the risky asset, a is the number of shares of the risky asset he purchases, and S is the number of shares of the safe asset he purchases. His expected utility (objective function) is

$$\max_{a,S} pu(S + aH) + (1-p)u(S + aL)$$

(i) Provide FONCs and SOSCs for the agent's investment problem; when are they satisfied? (ii) How does the optimal a^* vary with p and H? (iii) Can you figure out how the optimal a^* varies with wealth?

4. Re-write Theorem 5.2.1 to apply to minimization problems, rather than maximization problems, and provide a proof. Consider the consumer's *expenditure minimization problem*, $e(p, u) = \min_{q,m} pq + m$ subject to $v(q) + m = u$, where a consumer finds the cheapest bundle that provides $\bar{u}$ of utility. How does $q^*(p, u)$ vary in p and u?

5. Consider a representative consumer with utility function $u(q_1, q_2, m) = v(q_1, q_2) + m$ and budget constraint $w = p_1 q_1 + p_2 q_2 + m$. The two goods are produced by firm 1, whose costs are $C_1(q_1) = c_1 q_1$, and firm 2, whose costs are $C_2(q_2) = \dfrac{c_2}{2} q_2^2$. (i) Solve for the consumer's and firms' demand and supply curves. How does demand for good 1 change with respect to a change in the price of good 2? How does supply of good 1 change with respect to a change in the price of good 2? (ii) Solve for the system of equations that characterize equilibrium in the market. How does the equilibrium quantity of good 1 traded change with respect to a change in firm 2's marginal cost?

6. Do YOU want to be a New York Times columnist? Let's learn IS-LM in one exercise!

We start with a system of four equations: (i) The NIPA accounts equation,

$$Y = C + I + G$$

so that total national income, Y, equals total expenditure, consumption C plus investment I plus government expenditure G. (ii) The consumption function

$$C = f(Y - T)$$

giving consumption as a function of income less taxes. Assume $f'() \geq 0$, so that more income-less-taxes translates to more spending on consumption. (iii) The investment function

$$I = i(r)$$

where r is the interest rate. Assume $i'(r) \geq 0$, so that a higher interest rate leads to more investment. (iv) The money market equilibrium equation

$$M^s = h(Y, r)$$

so that supply of money must equal demand for money, which depends on national income and the interest rate. Assume $h_y \geq 0$, so that more income implies a higher demand for currency for trade, and $h_r \leq 0$, so that a higher interest rate moves resources away from consumption into investment, thereby reducing demand for currency.

This can be simplified to two equations,

$$Y - f(Y - T) - i(r) = G$$

$$h(Y, r) = M^s$$

The endogenous variables are Y and r, and the exogenous variables are α, β, γ, M^s, and G.

(i) What is the effect of increases in G affect Y and r? What does this suggest about fiscal interventions in the economy? (ii) How does an increase in M^s affect Y and r? What does this suggest about monetary interventions in the economy? (iii) Suppose there is a balanced-budget amendment so that $G = T$. How does an increase in G affect Y and r? Explain the effect of the amendment on the government's ability to affect the economy.

7. Consider a partial equilibrium model with a tax t where λ of the legal incidence falls on the consumer and $1 - \lambda$ falls on the firm. The consumer has objective $\max_q v(q) + w - (p + \lambda t)q$ and the firm has objective $\max_q (p - (1 - \lambda))q - C(q)$. (i) Compute the elasticity of the supply and demand curves with respect to p, λ, and t. (ii) Compute the change in the total taxes paid by the consumer and producer in equilibrium with respect to t and λ. (iii) Show that for a small increase in t starting from $t = 0$, the increase in the consumer's tax burden is larger than the producer's tax burden when the consumer's demand curve is more inelastic than the producer's supply curve.

Chapter 6

The Envelope Theorem

Recall our friend the monopolist:

$$\max_q p(q)q - \frac{c}{2}q^2$$

His FONCs are

$$p'(q^*)q^* - p(q^*) - cq^* = 0$$

and SOSCs are

$$p''(q^*)q^* - 2p'(q^*) - c < 0$$

This defines an implicit solution $q^*(c)$ in terms of the marginal cost parameter, c. But then we can define the *value function* or *indirect profit function* as

$$\pi(c) = p(q^*(c))q^*(c) - \frac{c}{2}(q^*(c))^2$$

giving the monopolist's maximized payoff for any value of c. If we differentiate with respect to c, we get

$$\pi'(c) = p'(q^*(c))\frac{\partial q^*(c)}{\partial c}q^*(c) + p(q^*(c))\frac{\partial q^*(c)}{\partial c} - \frac{1}{2}q^*(c)^2 - 2q^*(c)\frac{\partial q^*(c)}{\partial c}$$

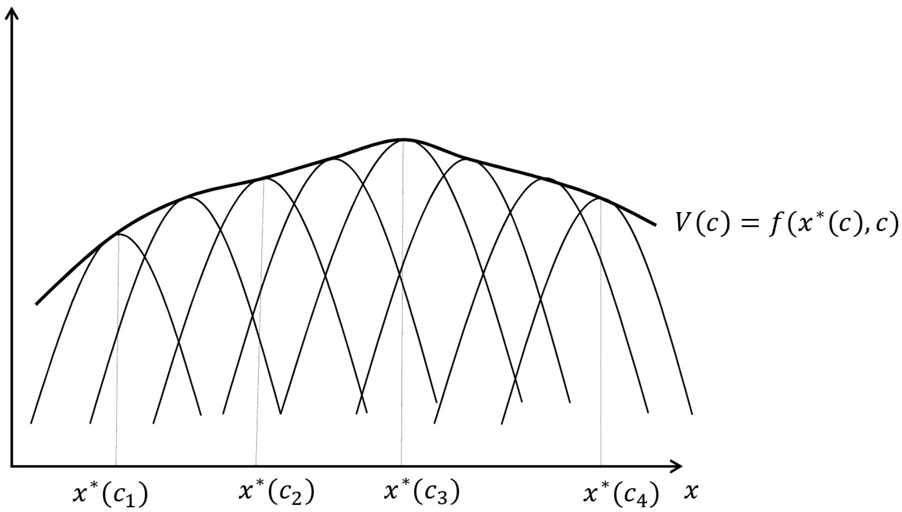
To sign this, it appears at first that we have to use the Implicit Function Theorem to derive $\partial q^*(c)/\partial c$, and then maybe substitute it in or just sign as much as we can. But wait! If we re-arrange it to put all the $\partial q^*(c)/\partial c$ terms together, we get

$$\pi'(c) = \underbrace{\left(p'(q^*(c))q^*(c) + p(q^*(c))2q^*(c)\right)}_{\text{FONCs}}\frac{\partial q^*(c)}{\partial c} - \frac{1}{2}q^*(c)^2$$

Since the FONCs are zero at $q^*(c)$, we get

$$\pi'(c) = -\frac{1}{2}q^*(c)^2 < 0$$

Without using the IFT at all. This is the basic idea of *Envelope Theorems*. The reason it is called an "envelope" is that the value function traces out a curve which is the maximum along all the global maxima of the objective function:



The Envelope Theorem

When you differentiate the value function, you are studying how the peaks shift.

6.1 The Envelope Theorem

Suppose an agent faces the maximization problem

$$\max_x f(x, c)$$

where c is some parameter. The FONC is

$$f_x(x^*(c), c) = 0$$

Now, consider the *value function* or *indirect payoff function*

$$V(c) = f(x^*(c), c)$$

This is the agent's optimized payoff, given the parameter c. We might want to know how $V(c)$ varies with c, or $V'(c)$. That derivative equals

$$V'(c) = \underbrace{f_x(x^*(c), c)}_{\text{FONC}} \frac{\partial x^*(c)}{\partial c} + f_c(x^*(c), c)$$

At first glance, it looks like we'll need to determine $\partial x^*(c)/\partial c$ using the implicit function theorem, substitute it in, and then try to sign the expression. But since $f_x(x^*(c), c)$ is zero at the optimum, the expression reduces to

$$V'(c) = f_c(x^*(c), c)$$

This means that the derivative of the value function with respect to a parameter is the *partial* derivative of the objective functions evaluated at the optimal solution.

Theorem 6.1.1 (Envelope Theorem) *For the maximization problem*

$$\max_x f(x, c)$$

the derivative of the value function is

$$V'(c) = \left[\frac{\partial f(x, c)}{\partial c}\right]_{x = x^*(c)}$$

Again, in words, the derivative of the value function is the *partial* derivative of $f(x, c)$ with respect to c, evaluated at the optimal choice, $x^*(c)$.

Example Consider

$$\pi(q, c) = pq - \frac{c}{2}q^2$$

The FONCs are

$$p - cq^*(p, c) = 0$$

Substituting this back into the objective yields

$$V(p, c) = \frac{p^2}{2c}$$

and

$$V_p(p, c) = \frac{p}{c} > 0$$

$$V_c(p, c) = -\frac{p^2}{2c^2} < 0$$

If we use the envelope theorem, however,

$$V_p(p, c) = \pi_p(q^*(p, c), p, c) = q^*(p, c) = \frac{p}{c} > 0$$

$$V_c(p, c) = \pi_c(q^*(p, c), p, c) = -\frac{1}{2}q^*(p, c)^2 = -\frac{p^2}{2c^2} < 0$$

These have the correct signs, and illustrate the usefulness of the envelope theorem. ∎

Example Suppose a consumer has utility function $u(q, m)$ and budget constraint $w = pq + m$. Then his value function will be

$$V(p, w) = u(q^*(p, w), w - pq^*(p, w))$$

Without even computing FONCs or SOSCs, then, we know that

$$V_p(p, w) = -u_m(q^*(p, w), w - pq^*(p, w))q^*(p, w)$$

and

$$V_w(p, w) = u_m(q^*(p, w), w - pq^*(p, w))$$

Do you see that without grinding through the FONCs? ∎

So we can figure out how changes in the environment affect the welfare of an agent without necessarily solving for the agent's optimal behavior explicitly. This can be very useful, not just in determining welfare changes, but also in simplifying the analysis of models. The next two examples illustrate how this works.

Example As a simple way of incorporating dynamic concerns into models, we might consider models of firms in which some variables are fixed — capital, technology, capacity, etc. — and maximize over a flexible control variable like price or quantity to generate a *short-run profit function.* Then, the fixed variable becomes flexible in the long-run, giving us the *long-run profit function.*

Suppose a price-taking firm's cost function depends on the quantity it produces, q, as well as its technology, t. Its short-run profit function is

$$\pi_S(q, p, t) = \max_q pq - C(q, t)$$

Let's stop now to think about what "better" technology means. Presumably, better technology should, at the least, mean that $C_t(q,t) < 0$, so that total cost is decreasing in t. Similarly, $C_{qt}(q,t) < 0$ would also make sense: Marginal cost $C_q(q,t)$ is decreasing in t. However, is this always the case? Some technologies have scale effects, where for $q < \bar{q}$, $C_{qt}(q,t) > 0$, but for $q > \bar{q}$, $C_{qt}(q,t) < 0$, so that the cost-reducing benefits are sensitive to scale. For example, having a high-tech factory where machines do all the labor will certainly be more efficient than an assembly line with workers, provided that enough cars are being made. Let's see what the FONCs and SOSCs we derive say about the relationships between these partials and cross-partials.

In the "short-run" we can treat t as fixed and study how the firm maximizes profits: It has an FONC

$$p - C_q(q,t) = 0$$

and SOSC

$$-C_{qq}(q,t) < 0$$

So as long as $C(q,t)$ is convex in q, there will be a unique profit-maximizing $q^*(t)$. Using the implicit function theorem, we can see how the optimal $q^*(t)$ varies in t:

$$-C_{qq}(q^*(p,t),t)\frac{\partial q^*(p,t)}{\partial t} - C_{tq}(q^*(p,t),t) = 0$$

or

$$\frac{\partial q^*(p,t)}{\partial t} = \frac{C_{tq}(q^*(t),t)}{-C_{qq}(q^*(t),t)}$$

If better technology lowers marginal cost, $C_{tq}(q,t) < 0$, and this $q^*(p,t)$ is increasing in t. Otherwise, $\partial q^*(p,t)/\partial t$ will be *decreasing* in t. Note that there is no contradiction between $C_t(q,t) < 0$ and $C_{qt}(q,t) > 0$.

But now suppose we take a step back to the "long-run". How should the firm invest in technology? Consider the long-run profit function

$$\pi_L(p,t) = \pi_S(q^*(p,t),p,t) - kt$$

where the $-kt$ term is the cost of adopting level t technology. Then the firm's FONC is

$$\frac{d}{dt}\left[\pi_S(q^*(p,t),p,t)\right] - k = 0$$

but by the envelope theorem, the first term is just the derivative of the short run value function with respect to t, so

$$-C_t(q^*(p,t^*(p)),t^*(p)) - k = 0$$

and the SOSC is

$$-C_{qt}(q^*(p,t^*(p)),t^*(p))\frac{\partial q^*(t)}{\partial t} - C_{tt}(q^*(p,t^*(p)),t^*(p)) < 0$$

Since $\partial q^*(p,t)/\partial t < 0$, for the SOSC to be satisfied it must be the case that $C_{tt}(q^*(p,t),t) < 0$. This would mean that the marginal cost reduction in t is decreasing in t: Better technology always reduces the total cost function, $C_t(q,t) < 0$, but at a decreasing rate $C_{tt}(q,t) < 0$.

How does the profit-maximizing $t^*(p)$ depend on price? Using the implicit function theorem,

$$-C_{qt}(q^*(p,t^*(p)),t^*(p))\left(\frac{\partial q^*(p,t^*(p))}{\partial p} + \frac{\partial q^*(p,t^*(p))}{\partial t}\frac{\partial t^*(p)}{\partial p}\right) - C_{tt}(q^*(p,t^*(p)),t^*(p))\frac{\partial t^*(p)}{\partial p} = 0$$

or

$$\frac{\partial t^*}{\partial p} = \frac{C_{qt}\dfrac{\partial q^*}{\partial p}}{-C_{qt}\dfrac{\partial q^*}{\partial t} - C_{tt}} = \frac{\underbrace{\dfrac{\partial q^*}{\partial p}}_{+}}{-\dfrac{\partial q^*}{\partial t} - C_{tt}/C_{qt}}$$

The numerator is positive, and the denominator is positive if (using the comparative static from the short-run problem)

$$-\frac{C_{tq}}{-C_{qq}} - C_{tt}/C_{qt} > 0$$

or

$$C_{qt}C_{tq} - C_{qq}C_{tt} > 0$$

It will turn out that the above inequality implies that $C(q,t)$ is a convex function when considered as a two-dimensional object.

This kind of short-run/long-run model is very useful in providing simple but rigorous models of how firms behave across time. Of course, there are no dynamics here, but it captures the idea of how in the short run the firm can vary output but not technology, but in the long run things like technology, capital, and other "investment goods" become choice variables. ∎

Example This is an advanced example of how outrageously powerful the envelope theorem can be. In particular, we'll use it to derive the (Bayesian Nash) equilibrium strategies of a first-price auction.

At a first-price auction, there are $i = 1, 2, ..., N$ buyers competing for a good. Buyer i knows his own value, $v_i > 0$, but no one else's. The buyers simultaneously submit a bid b_i. The highest bidder wins, and gets a payoff $v_i - b_i$, while the losers get nothing. Let $p(b_i)$ be the probability that i wins given a bid of b_i.

Presumably, all the buyers' bids should be increasing in their values. For example, if buyer 1's value is 5 and buyer 2's value is 3, buyer 1 should bid a higher amount. Said another way, the bid function $b_i = b(v_i)$ is increasing. Then buyer i's expected payoff is

$$U(v_i) = \max_{b_i} p(b_i)(v_i - b_i)$$

with FONC

$$p'(b_i)(v_i - b_i) - p(b_i) = 0$$

The envelope theorem implies

$$U'(v_i) = p(b(v_i))$$

And integrating with respect to v_i yields

$$U(v_i) = \int_0^{v_i} p(b(x))dx$$

Now, equating the two expressions for $U(v_i)$ gives

$$p(b(v_i))(v_i - b(v_i)) = U(v_i) = \int_0^{v_i} p(b(x))dx$$

and re-arranging yields

$$b(v) = v - \frac{\int_0^v p(b(x))dx}{p(b(v))}$$

If we knew the probability that an agent with value v making a bid $b(v)$ won, we could solve the about expression. But consider the rules of the auction: The highest bidder wins. If $b(v)$ is increasing, then the probability of i being the highest bidder with bid $b(v_i)$ is

$$\begin{aligned} &pr[b(v_i) \geq b(v_1), ..., b(v_i) \geq b(v_{i-1}), b(v_i) \geq b(v_{i+1}), ..., b(v_i) \geq b(v_N)] \\ &\qquad = pr[v_i \geq v_1, ..., v_i \geq v_{i-1}, v_i \geq v_{i+1}, ..., v_i \geq v_N] \end{aligned}$$

If the bidders' values are independent and the probability distribution of each bidder's value v is $F(v)$, then the probability of having the highest value given v_i is

$$\begin{aligned} &pr[v_i \geq v_1, ..., v_i \geq v_{i-1}, v_i \geq v_{i+1}, ..., v_i \geq v_N] \\ &\qquad = pr[v_i \geq v_1]...pr[v_i \geq v_{i-1}]pr[v_i \geq v_{i+1}]...pr[v_i \geq v_N] \\ &\qquad = F(v_i)...F(v_i)F(v_i)...F(v_i) = F(v_i)^{N-1} \end{aligned}$$

So that

$$p(b(v_i)) = F(v_i)^{N-1}$$

and

$$b(v) = v - \frac{\int_0^v F(x)^{N-1}dx}{F(v)^{N-1}}$$

Since this function is indeed increasing, it satisfies the FONC above. So using the envelope theorem and some basic probability, we've just derived the strategies in one of the most important strategic games that economists study. Other derivations require solving systems of differential equations or using a sub-field of game theory called mechanism design. $\blacksquare$

Exercises

1. Suppose that a price-taking firm can vary its choice of labor, L, in the short run, but capital, K is fixed. Quantity q is produced using technology $F(K, L)$. In the long run the firm can vary capital. The cost of labor is w, and the cost of capital is r, and the price of its good is p. (i) Derive the short-run profit function and show how q^* varies with r and p. (ii) Derive the long-run profit function and show how K varies with r. (iii) How does a change in p affect the long-run choices of K and L, and the short run choice of L?

2. Suppose an agent maximizes

$$\max_x f(x, c)$$

yielding a solution $x * (c)$. Suppose the parameter c is perturbed to c'. Use a second-order Taylor series expansion to characterize the loss that arises from using $x^*(c)$ instead of the new maximizer, $x^*(c')$. Show that the loss is proportional to the square of the maximization error, $x^*(c') - x^*(c)$.

3. Suppose consumers have utility function $u(Q, m) = b\log(Q) + m$ and face a budget constraint $w = pQ + m$. Solve for the firms' short-run profit functions in equilibrium. (i) If there are K price-taking firms, each with cost function $c(q) = \dfrac{c}{2}q^2$ and aggregate supply is $Q = Kq$, solve for the firms' short-run profit functions $\pi_S(K)$ in terms of K. (ii) If there is a fixed cost F to entry and there no other barriers to entry, characterize the long-run profit function $\pi_L(K)$ and solve for the long-run number of firms K^*. How does K^* vary in b, c, and F? (iii) Can you generalize this analysis to a convex, increasing $C(q)$ and concave, increasing $v(q)$ using the IFT and envelope theorem? In particular, how do p^* and q^* respond to an increase in F, and how does K^* respond to an increase in F?

4. Consider a partial equilibrium model with a consumer with utility function $u(q, m) = v(q) + m$ and budget constraint $w = pq + m$, and a firm with cost function $C(q)$. Suppose the firm must pay a tax t for each unit of the good it sells. Let social welfare be given by

$$W(t) = (v(q^*(t)) - p^*(t)q^*(t) + w) + ((p^*(t) - t)q^*(t) - C(q^*(t))) + tq^*(t)$$

Use a second-order Taylor series to approximate the loss in welfare from the tax, $W(t) - W(0)$. Show that this welfare loss is approximately proportional to the square of the tax. Sketch a graph of the situation.

5. Consider a partial equilibrium model with a consumer with utility function $u(q, m, R) = v(q) + m + g(R)$ and budget constraint $w = pq + m$, where R is government expenditure and $g(R)$ is the benefit to the consumers from government expenditures. The firm has a cost function $C(q)$, and pays a tax t for each unit of the good it sells. Tax revenue is used to fund government services, which yield benefit $g(R)$ to consumers, where $R = tq$ is total tax revenue. Define *social welfare* as

$$W(t) = (v(q) - pq + w) + ((p - t)q - C(q)) + g(tq)$$

where the last term represents tax revenue. (i) If the firm and consumer maximize their private benefit taking R as a constant, what are the market-clearing price and quantity? How do the market-clearing price and quantity vary with the tax? (ii) What is the welfare maximizing level of t?

Chapter 7

Concavity and Convexity

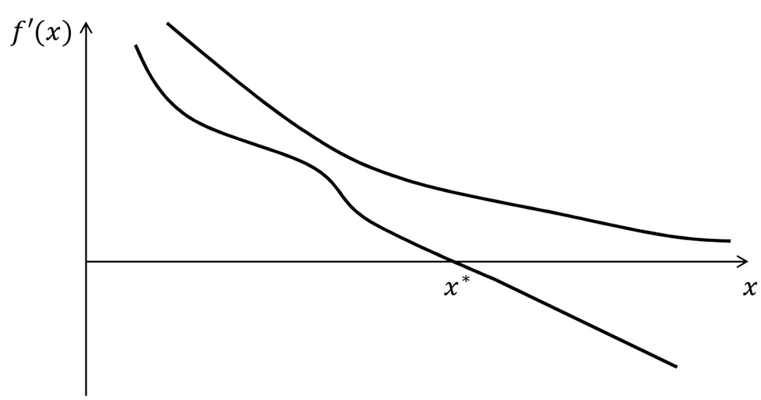
At various times, we've seen functions for which the SOSCs are satisfied for any point, not just the critical point. For example,

- A price-taking firm with cost function $C(q) = \dfrac{c}{2}q^2$ has SOSCs $-c < 0$, independent of q
- A price-taking consumer with benefit function $v(q) = b\log(q)$ has SOSCs $-1/q^2 < 0$, which is satisfied for any q
- A household with savings problem $\max_s \log(y_1 - s) + \delta \log(Rs + y_2)$ has SOSCs
$$\frac{-1}{(y_1 - s)^2} + \delta R^2 \frac{-1}{(Rs + y_2)^2} < 0$$
which is satisfied for any s

The common feature that ties all these examples together is that

$$f''(x) < 0$$

for *all* x, not just a critical point x^*. This implies that the first-order necessary condition $f'(x)$ is a monotone decreasing function in x, so if it has a zero, it can only have one (keep in mind though, some decreasing functions have no zeros, like e^{-x}).



Concave functions have a unique critical point, or none

This actually solves all our problems with identifying maxima: If a function satisfies $f''(x) < 0$, it has at most one critical point, and any critical point will be a global maximum. This is a special class of functions that deserves some study.

7.1 Concavity

Recall the partial equilibrium consumer with preferences $u(q,m) = v(q) + m$. Earlier, we claimed that $v'(q) > 0$ and $v''(q) < 0$ were good economic assumptions: The agent has positive marginal value for each additional unit, but this marginal value is decreasing. Let's see what these economic assumptions imply about Taylor series. The benefit function, $v(q)$, satisfies

$$v(q) = v(q_0) + v'(q_0)(q - q_0) + v''(q_0)\frac{(q-q_0)^2}{2} + o(h^3)$$

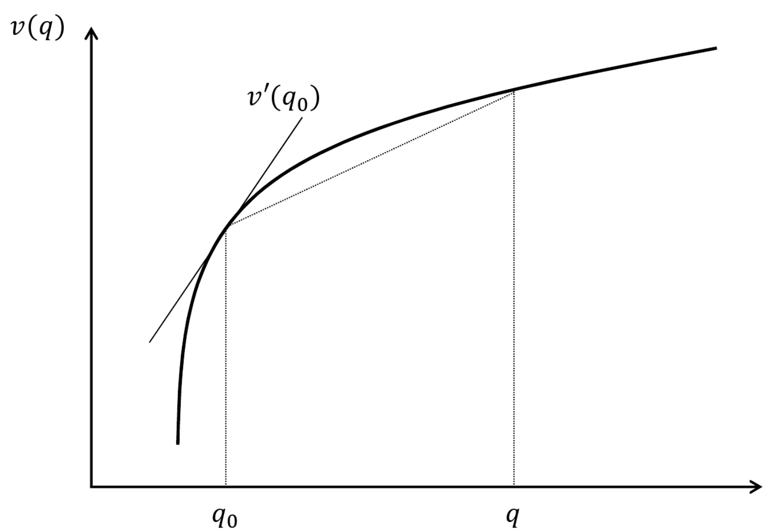
If we re-arrange this a little bit, we get

$$\frac{v(q) - v(q_0)}{q - q_0} = v'(q_0) + v''(q_0)\frac{q - q_0}{2} + o(h^3)$$

and since we know that $v''(q_0) \leq 0$, this implies, for q_0 close to q

$$\frac{v(q) - v(q_0)}{q - q_0} \leq v'(q_0)$$

In words or pictures, the derivative at q_0 is steeper than the chord from $v(q_0)$ to $v(q)$.



Concavity

This is one way of defining the idea of a *concave function*. Here are the standard, equivalent definitions:

Theorem 7.1.1 *The following are equivalent: Let $f : D = (a,b) \to \mathbb{R}$.*

- *$f(x)$ is* concave on D
- *For all $\lambda \in (0,1)$ and x', x'' in D, $f(\lambda x' + (1-\lambda)x'') \geq \lambda f(x') + (1-\lambda)f(x'')$*
- *$f''(x) \leq 0$ for all x in D, where $f()$ twice differentiable*
- *For all x' and x'',*

$$f'(x') \geq \frac{f(x'') - f(x')}{x'' - x'}$$

If the weak inequalities above hold as strict inequalities, then $f(x)$ is strictly concave.

Note that concavity is a global property (holds on all of D) that depends on the domain of the function, $D = (a, b)$. For example, $\log(x)$ is concave on $(0, \infty)$, since

$$f''(x) = \frac{-1}{x^2} < 0$$

for *all* x in $(0, \infty)$. The function $\sqrt{|x|}$, however, is concave for $(0, \infty)$, and concave for $(-\infty, 0)$, but not concave for $(-\infty, \infty)$. Why? If we connect the points at $\sqrt{|\pm 1|} = 1$ by a chord, it lies above the function at $\sqrt{|0|} = 0$, violating the second criterion for concavity. So a function can have $f''(x) \leq 0$ for *some* x and not be concave: It is concave only if $f''(x) < 0$ for *all* x in its domain, D.

Theorem 7.1.2 *If $f(x)$ is differentiable and strictly concave, then it has at most one critical point which is the unique global maximum.*

Proof If $f(x)$ is strictly concave, the FONC is

$$f'(x) = 0$$

Since $f(x)$ is strictly concave, the derivative $f'(x)$ is monotone decreasing. This implies there is at most one time that $f'(x)$ crosses zero, so it has at most one critical point. Since

$$f(x^*) - f(x) = -f''(x^*)\frac{(x - x^*)^2}{2} > 0$$

holds (if $f''(x) < 0$ for all x, not just x^*), this is a local maximum. Since there is a unique critical point x^* and $f''(x) < 0$ for all x, the candidate list consists only of x^*. ∎

There is some room for confusion about what the previous theorem buys you, because a concave function might not have any critical points. For example,

$$f(x) = \log(x)$$

has FONC

$$f'(x) = \frac{1}{x}$$

which reaches $f'(x) = 0$ only as $x \to \infty$ (∞ is not a number). Since $\log(x) \to$ as $x \to \infty$, there isn't a finite maximizer. On the other hand

$$f(x) = \log(x) - x$$

has FONC

$$f'(x) = \frac{1}{x} - 1$$

and SOSC

$$f''(x) = -\frac{1}{x^2}$$

so that $x^* = 1$ is the unique global maximizer.

7.2 Convexity

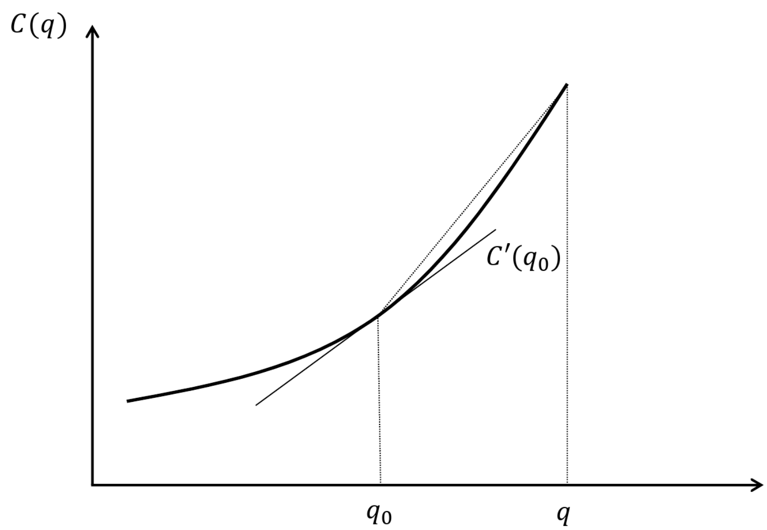
Similarly, our cost function $C(q)$, satisfies

$$C(q) = C(q_0) + C'(q_0)(q - q_0) + C''(c)\frac{(q - q_0)^2}{2}$$

Re-arranging it the same way, we get the opposite conclusion,

$$\frac{C(q) - C(q_0)}{q - q_0} \geq C'(q_0)$$

so that the chord from $C(q_0)$ to $C(q)$ is steeper than the derivative $C'(q_0)$. This is a *convex* function.



Convexity

Theorem 7.2.1 *The following are equivalent: Let $f : D = (a, b) \to \mathbb{R}$.*

- *$f(x)$ is a* convex *on D.*
- *For all $\lambda \in (0, 1)$ and x', x'' in D, $f(\lambda x' + (1 - \lambda)x'') \leq \lambda f(x') + (1 - \lambda)f(x'')$*
- *$f''(x) > 0$ for all x in D*
- *If $f(x') \geq f(x''$ for all x', x'' in D, then*

$$\frac{f(x'') - f(x')}{x'' - x'} \geq f'(x')$$

If the above weak inequalities hold strictly, then $f(x)$ is strictly convex*.*

Convexity is a useful property for, in particular, minimization.

Exercises

1. Rewrite Theorem 7.1.2, replacing the word "concave" with "convex" and "maximization" with "minimization." Sketch a graph like the first one in the chapter to illustrate the proof.

2. Show that if $f(x)$ is concave, then $-f(x)$ is convex. If $f(x)$ is convex, then $-f(x)$ is concave.

3. Show that if $f_1(x), f_2(x), ..., f_n(x)$ are concave, then $f(x) = \sum_{i=1}^{n} f_i(x)$ is concave. If $f_1(x), f_2(x), ..., f_n(x)$ are convex, then $f(x) = \sum_{i=1}^{n} f_i(x)$ is convex.

4. Can concave or convex functions have discontinuities? Provide an example of a concave or convex function with a discontinuity, or show why they must be continuous.

Part II

Optimization and Comparative Statics in $\mathbb{R}^N$

Chapter 8

Basics of $\mathbb{R}^N$

Now we need to generalize all the key results (extreme value theorem, FONCs/SOSCs, and the implicit function theorem) to situations in which many decisions are being made at once, sometimes subject to constraints. Since we need functions involving many choice variables and possibly many parameters, we need to generalize the real numbers to $\mathbb{R}^N$, or *N-dimensional Euclidean vector space.*

An *N-dimensional vector* is a list of N ordered real numbers,

$$x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{pmatrix}$$

where certain rules of addition and multiplication are defined. In particular, if x and y are vectors and a is a real number, *scalar multiplication* is defined as

$$ax = \begin{pmatrix} ax_1 \\ ax_2 \\ \vdots \\ ax_N \end{pmatrix}$$

and *vector addition* is defined as

$$x + y = \begin{pmatrix} x_1 + y_1 \\ x_2 + y_2 \\ \vdots \\ x_N + y_N \end{pmatrix}$$

The *transpose* of a column vector is a row vector,

$$x' = (x_1, x_2, ..., x_N)$$

and vice versa. A *basis vector* e_i is a vector with a 1 in the i-th spot, and zeros everywhere else.

The set of N-dimensional vectors with real entries $\mathbb{R}^N$ is *Euclidean*, since we define length by the *Euclidean norm.* The norm is a generalization of absolute value, and is given by

$$||x|| = \sqrt{x_1^2 + x_2^2 + ... + x_N^2}$$

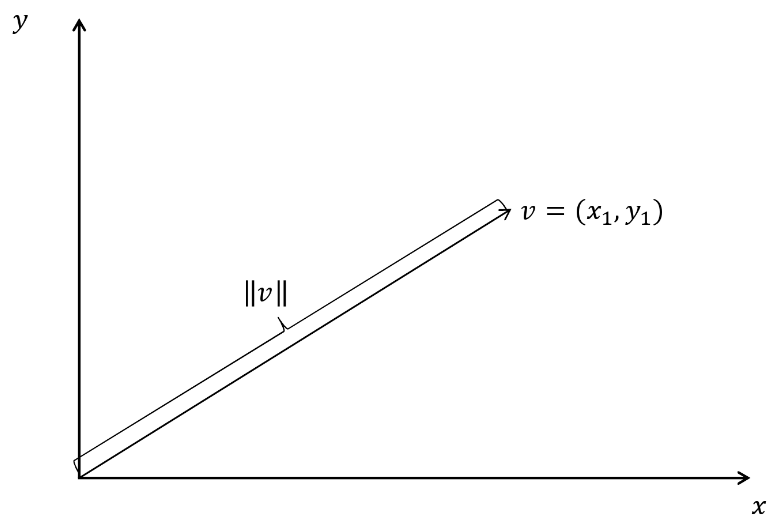
Though, we could just as easily use

$$||x|| = \max_i x_i$$

or

$$||x|| = \left(\sum_i x_i^p\right)^{1/p}$$

These are all just ways of summarizing how “large” a vector is.



The Norm

For $N = 1$, this is just $\mathbb{R}$, since $||x|| = \sqrt{x^2} = |x|$. With this norm in mind, we define *distance* between two vectors x and y as

$$||x - y|| = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + ... + (x_N - y_N)^2}$$

Instead of defining an open interval as (a, b), we define the *open ball of radius δ at x_0*, given by

$$B_\delta(x_0) = \{y : ||y - x_0|| < \delta\}$$

and instead of defining a closed interval $[a, b]$, we define the *closed ball of radius δ at x_0* is

$$\bar{B}_\delta(x_0) = \{y : ||y - x_0|| \leq \delta\}$$

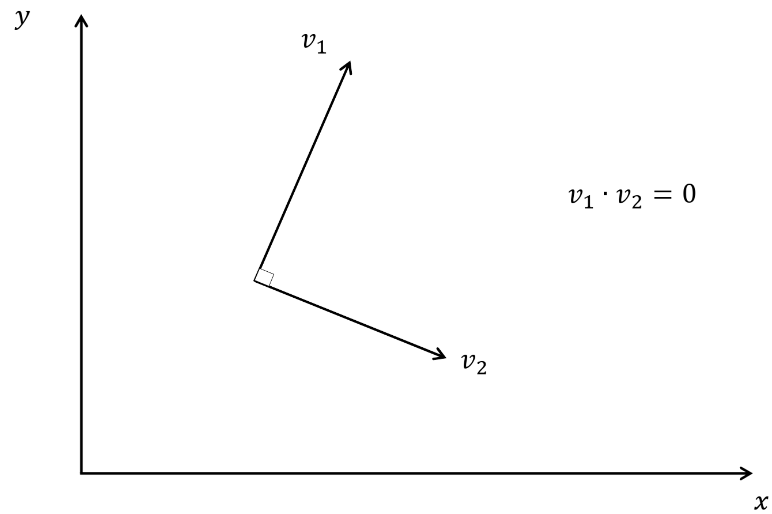
The last special piece of structure on Euclidean space is *vector multiplication*, or the *inner product* or *dot product*:

$$x'y =< x, y >= x \cdot y = x_1y_1 + x_2y_2 + ... + x_Ny_N$$

In terms of geometric intuition, the dot product is related to the angle θ between x and y:

$$\cos(\theta) = \frac{x \cdot y}{||x||||y||}$$

Note that if $x \cdot y = 0$, then $\cos(\theta) = 0$, so that the vectors must be at right angles to each other, or *x is orthogonal to y.*



Orthogonal Vectors

So a Euclidean space has notions of direction, distance, and angle, and open and closed balls are easy to define. Not all spaces have these properties, which makes Euclidean space special[1].

8.1 Intervals $\rightarrow$ Topology

In $\mathbb{R}$, we had open sets (a, b), closed sets $[a, b]$, and bounded sets, where a and b are finite. In multiple dimensions, however, it's not immediately obvious what the generalizations of these properties should be. For example, we could define *open cells* as sets $C = (0, 1) \times (0, 1) \times ...(0, 1)$, which look like squares or boxes. Or we could define *open balls* as sets $B = \{y : ||y|| < 1\}$. But then, we're going to study things like budget sets, defined by inequalities like $\{(x, y) : x \geq 0, y \geq 0, p_x x + p_y y \leq w\}$, which are neither balls nor cells.

To avoid these difficulties, we'll use some ideas from topology. Topology is the study of properties of sets, and what kinds of functions preserve those properties. For example, the function $f(x) = x$ maps the set $[-1, 1]$ into the set $[-1, 1]$, but the function

$$g(x) = \begin{cases} x & \text{if } |x| \neq 1 \\ 0 & \text{if } |x| = 1 \end{cases}$$

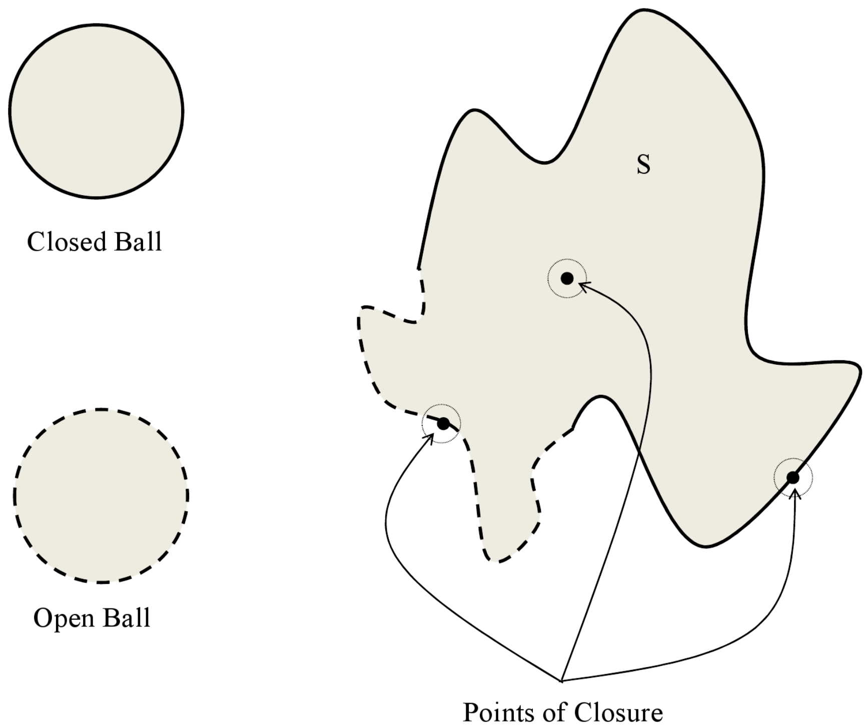
maps $[-1, 1]$ into $(-1, 1)$. So $f(x)$ maps a closed set to a closed set, while $g(x)$ maps a closed set to an open set. So the question, "What kinds of functions map closed sets to closed sets?" is a topological question.

Why do YOU care about topology? What we want to do is study the image of a function, $f(D)$ and see if it achieves its maximum. The function $f(x) = x$ above achieves a maximum at $x = 1$. The second function, $g(x)$, does not achieve a maximum, since $\sup g([-1, 1]) = 1$ but 1 is not in $g([-1, 1])$. Our one-dimensional Weierstrass theorem tells us that we shouldn't expect $g(x)$ to achieve a maximum since it is not a continuous function. But how do we generalize this to Euclidean space and more general sets than simple intervals (right? Because now we might be maximizing over spheres, triangles, tetrahedrons, simplexes, and all kinds of non-pathological sets that are more complicated than $[a, b]$)?

[1] For example, the space of continuous functions over a closed interval $[a, b]$ is also a vector space, called $C([a, b])$, with scalar multiplication $fa = (af(x))$, vector addition $f + g = (f(x) + g(x))$, norm $||f|| = \max_{x \in [a,b]} f(x)$ and distance $d(f, g) = \max_{x \in [a,b]} |f(x) - g(x)|$. You use this space all the time when you do dynamic programming in macroeconomics, since you are looking for an unknown function that satisfies certain properties. This space has very different properties — for example, the closed unit ball $\{g : d(f, g) \leq 1\}$ is not compact.

Definition 8.1.1 • *A set is* open *if it is a union of open balls*

- x *is a* point of closure of a set S *if for every* $\delta > 0$*, there is a* y *in* S *such that* $||x - y|| < \delta$
- *The set of all points of closure of a set* S *is* $\bar{S}$*, and* $\bar{S}$ *is the* closure of S
- *A set* S *is* closed *if* $\bar{S} = S$
- *A set* S *is* closed *if its complement,* S^c *is open*
- *A set* S *is* bounded *if it is a subset of an open ball* $B_\delta(x_0)$*, where* δ *is finite. Otherwise, the set is* unbounded



Open Balls, Closed Balls, and Points of Closure

Note that x can be a point of closure of a set S without being a member of S. For example, the open interval $(-1, 1)$ is an open set, since it is an open ball, $B_1(0) = \{y : |y - 0| < 1\} = (-1, 1)$. Each point in $B_1(0)$ is a point of closure of $(-1, 1)$, since for all $\delta > 0$ they contain an element of $(-1, 1)$. However, the points -1 and 1 are also points of closure of $(-1, 1)$, since for all $\delta > 0$, -1 and 1 contain some points of $(-1, 1)$. So the closure of $(-1, 1)$ is $[-1, 1]$. As you might have expected, this is a closed set.

This works just as well in Euclidean space. The open ball at zero, $B_\delta(0) = \{y : ||y - 0|| < \delta\}$ is an open ball, so it is a union of open balls, so it is open. All elements of $B_\delta(0)$ are points of closure of $B_\delta(0)$, since for all $\delta > 0$, they contain a point of $B_\delta(0)$. However, the points satisfying $\{y : ||y - 0|| = \delta\}$ are also points of closure, since part of the open ball around any of these points must intersect with $B_\delta(0)$. Consequently, the set of all points of closure of $B_\delta(0)$ is the set $\{y : ||y - 0|| \leq \delta\}$, which is the closed ball.

There are many other properties that are considered topological: connected, dense, convex, separable, and so on. But for our purposes of developing maximization theory, there is a special topological property that will buy us what we need to generalize the extreme value theorem:

Definition 8.1.2 *A subset K of Euclidean space is* compact *if each sequence in K has a subsequence that converges to a point in K.*

The Bolzano-Weierstrass theorem (which is a statement about the properties of sequences) looks similar, but this definition concerns the properties of a sets. A set is called compact if any sequence constructed from it has the Balzano-Weierstrass property. For example, the set (a, b) is not compact, because $x_n = b - 1/n$ is a sequence entirely in (a, b), but all of its subsequences converge to b which is not in (a, b). Characterizing compact sets in a space is the typical starting point for studying optimization in that space[2]. As it happens, there is an easy characterization for $\mathbb{R}^N$:

Theorem 8.1.3 (Heine-Borel) *In Euclidean space, a set is* compact *iff it is closed and bounded.*

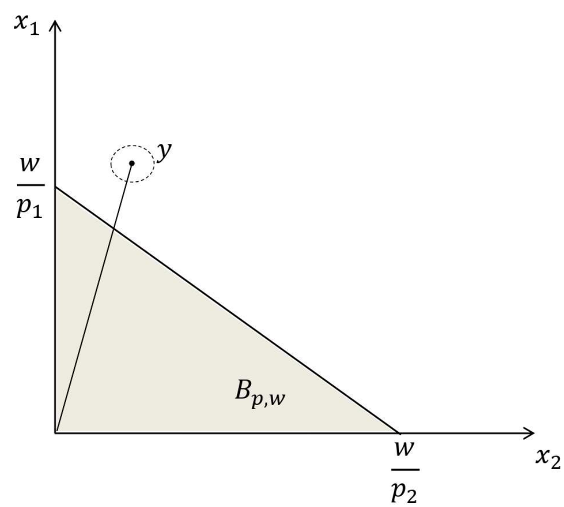
Basically, bounded sets that include all their points of closure are compact in $\mathbb{R}^N$. Non-compact sets are either unbounded, like $\{(x, y) : (x, y) \geq (0, 0)\}$, or open, like $\{(x, y) : x^2 + y^2 < c\}$.

Example Consider the set in $\mathbb{R}^N$ described by $x_i \geq 0$ for all $i = 1, ..., N$, and

$$\sum_{i=1}^{N} p_i x_i = p \cdot x \leq w$$

This is called a budget set, $B_{p,w}$. In two dimensions, this looks like a triangle with vertices at 0, w/p_1 and w/p_2. This set is compact.

We'll use the Heine-Borel theorem, and show the set is closed and bounded.



Budget Sets

It is bounded, since if we take $\delta = \max_i w/p_i + 1$, the set is contained in $B_\delta(0)$ (right?). It is closed since if we take any sequence x_n satisfying $x_n \geq 0$ for all n and $p \cdot x_n \leq w$ for all n, the limit of the sequence must satisfy these inequalities as well (a weakly positive sequence for all terms can't become negative at the limit). If you don't like that argument that it is closed, we can prove $B_{p,w}$ is closed by showing that the complement of $B_{p,w}$ is open: Let y be a point outside $B_{p,w}$. Then the ray $z = \alpha 0 + (1 - \alpha)y$ where $\alpha \in [0, 1]$ starts in the budget set, but eventually exits it to reach y. Take α^* to be the point at which the ray exits the set. Then we can draw an open ball around y with radius $r = (1 - \alpha^*)/2$, so that $B_r(y)$ contains no points of $B_{p,w}$. That means that for *any* y, we can draw an open ball around it that contains no point in $B_{p,w}$. That means that $B_{p,w}^c$ is a union of open balls, so $B_{p,w}$ is closed. ∎

[2]See the footnote on page 1: Closed, bounded sets in $C([a, b])$ are not compact.

So competitive budget sets are compact. In fact, we've shown that any set characterized by $x \geq 0$ and $a \cdot x \leq c$ is compact, which is actually a sizeable number of sets of interest.

Example The *N-dimensional unit simplex*, Δ_N is constructed by taking the N basis vectors, $(e_1, e_2, ..., e_n)$, and considering all *convex combinations*

$$x_\lambda = \lambda_1 e_1 + \lambda_2 e_2 + ... + \lambda_N e_N$$

such that $\sum_{i=1}^{N} \lambda_i = 1$ and $0 \leq \lambda_i \leq 1$.

Then Δ_N is bounded, since if we take the N-dimensional open ball $B_2(0)$ of radius 2, it includes the simple. And Δ_N is closed, since, taking any sequence $\lambda_n \to \lambda$, and $\lim_{n\to\infty} \lambda_n'(e_1, ..., e_N) = \lambda'(e_1, ..., e_N)$, which is in the simplex. So the simplex is its own closure, $\bar{\Delta}_N = \Delta_N$, so it is closed. $\blacksquare$

The simplex comes up in probability theory all the time: It is the set of all probability distributions over N outcomes.

8.2 Continuity

Having generalized the idea of "$[a, b]$" to $\mathbb{R}^N$, we now need to generalize continuity. Continuity is more difficult to visualize in $\mathbb{R}^N$ since we can no longer sketch a graph on a piece of paper. For a function $f : \mathbb{R}^2 \to \mathbb{R}$ we can visualize the graph in three dimensions. A continuous function is one for which the "sheet" is relatively smooth: It may have "ridges" or "kinks" like a crumpled-up piece of paper that has been smoothed out, but there are no "rips" or "tears" in the surface.

Definition 8.2.1 *A function $f : D \to \mathbb{R}$ is* continuous *at $c \in D$ if for all $\varepsilon > 0$, there exists a $\delta > 0$ so that if $||x - c|| < \delta$, then $|f(x) - f(c)| < \varepsilon$.*

The only modification from the one-dimensional definition is that we have replaced the set $(c - \delta, c + \delta)$ with an open ball, $B_\delta(c) = \{x : ||x - c|| < \delta\}$. Otherwise, everything is the same. Subsequently, the proof of the Sequential Criterion for Continuity is almost exactly the same:

Theorem 8.2.2 (Sequential Criterion for Continuity) *A function $f : D \to \mathbb{R}$ is continuous at c iff for all sequences $x_n \to c$, we have*

$$\lim_{n\to\infty} f(x_n) = f\left(\lim_{n\to\infty} x_n\right) = f(c)$$

Again, continuity allows us to commute the function $f()$ with a limit operator lim. This is the last piece of the puzzle of generalizing the extreme value theorem.

8.3 The Extreme Value Theorem

So, what do you have to internalize from the preceding discussion in this chapter?

- A set is open if it is a union of open balls.
- A set is closed if it contains all its points of closure. A set is closed if its complement is open.
- A set is compact in $\mathbb{R}^N$ if it is closed and bounded.
- In compact sets, all sequences have a convergent subsequence.

- If a function is continuous in $\mathbb{R}^N$, then

$$\lim_{n \to \infty} f(x_n) = f\left(\lim_{x \to \infty} x_n\right)$$

for any sequence $x_n \to c$.

The above facts let us generalize the Extreme Value Theorem:

Theorem 8.3.1 (Weierstrass' Extreme Value Theorem) *If K is a compact subset of $\mathbb{R}^N$ and $f : K \to \mathbb{R}$ is a continuous function, then $f(x)$ achieves a maximum on K.*

Proof Let

$$\sup f(K) = m^*$$

Then we can construct a sequence x_n satisfying

$$m^* - \frac{1}{n} \le f(x_n) \le m^*$$

Since K is compact, every sequence x_n has a convergent subsequence $x_{n_k} \to x^*$. Taking limits throughout the inequality, we get

$$\lim_{n_k \to \infty} m^* - \frac{1}{n_k} \le \lim_{n_k \to \infty} f(x_{n_k}) \le m^*$$

and by continuity of $f()$,

$$\lim_{n_k \to \infty} m^* - \frac{1}{n_k} \le f\left(\lim_{n_k \to \infty} x_{n_k}\right) \le m^*$$

so that

$$m^* \le f(x^*) \le m^*$$

Since K is closed, the limit x^* of the convergent subsequence x_{n_k} is in K (See the proof appendix; all closed sets contain all of their limit points).

Therefore, a maximizer exists since x^* is in K and achieves the supremum, $f(x^*) = m^*$. ■

This is the key result in explaining when maximizers and minimizers exist in $\mathbb{R}^N$. Moreover, the assumptions are generally easy to verify: Compact sets are closed and bounded, which we can easily find in Euclidean space, and since our maximization problems generally involve calculus, $f(x)$ will usually be differentiable, which implies continuity.

8.4 Multivariate Calculus

As we discussed earlier, differentiability of a function $f : \mathbb{R}^N \to \mathbb{R}$ is slightly more complicated that for the one-dimensional case. We review some of that material here briefly and introduce some new concepts.

Definition 8.4.1

- *The* partial derivative of $f(x)$ with respect to x_i*, where $f : D \to \mathbb{R}$ and D is an open subset of $\mathbb{R}$, is*

$$\frac{\partial f(x)}{\partial x_i} = f_{x_i}(x) = \lim_{h \to 0} \frac{f(x_1, x_2, ..., x_{i-1}, x_i + h, x_{i+1}, ..., x_N) - f(x_1, x_2, ..., x_{i-1}, x_i, x_{i+1}, ..., x_N)}{h}$$

- *The* gradient of $f(x)$ *is the vector of partial derivatives,*

$$\nabla f(x) = \left(\frac{\partial f(x)}{\partial x_1}, \frac{\partial f(x)}{\partial x_2}, ..., \frac{\partial f(x)}{\partial x_N}\right)$$

- *The* total differential of $f(x)$ *at* x *is*

$$df(x) = \sum_{i=1}^{N} f_{x_i}(x) dx_i$$

Now that we have a slightly better understanding of $\mathbb{R}^N$, the geometric intuition of the gradient is more clear. The partial derivative with respect to x_i is a one-dimensional derivative in the i-th dimension, giving the change in the function value that can be attributed to perturbing x_i slightly. The gradient is just the vector from the point $(x, f(x))$ pointing in the direction $\nabla f(x)$, which represents how the function is changing at that point.

If we multiple $\nabla f(x)$ by the basis vector e_i — where $e_i = 1$ in the i-th entry, but is zero otherwise — we get

$$\nabla f(x) \cdot e_i = \frac{\partial f(x)}{\partial x_i}$$

so we are asking, "If we increase x_i a small amount, how does $f(x)$ change?" What if we wanted to investigate how $f(x)$ changes in some other direction, y? The change in $f(x)$ in the direction yh as h goes to zero is given by

$$\lim_{h \to 0} \frac{f(x_1 + y_1 h, ..., x_N + y_N h) - f(x_1, ...x_N)}{h}$$

To compute this, we can use a Taylor series dimension-by-dimension on $f(x_1 + y_1 h, x_2 + y_2 h)$ to get

$$f(x_1 + y_1 h, x_2 + y_2 h) = f(x_1, x_2 + y_2 h) + \frac{\partial f(\xi_1, x_2 + y_2 h)}{\partial x_1} y_1 h$$

where ξ_1 is between x_1 and $x_1 + y_1 h$. Doing this again with respect to x_2 gives

$$f(x_1, x_2 + y_2 h) = f(x_1, x_2) + \frac{\partial f(x_1, \xi_2)}{\partial x_2} y_2 h$$

Substituting the above equation into the previous one yields

$$f(x_1 + y_1 h, x_2 + y_2 h) = f(x_1, x_2) + \frac{\partial f(x_1, \xi_2)}{\partial x_2} y_2 h + \frac{\partial f(\xi_1, x_2 + y_2 h)}{\partial x_1} y_1 h$$

Re-arranging and dividing by h yields

$$\frac{f(x_1 + y_1 h, x_2 + y_2 h) - f(x_1, x_2)}{h} = \frac{\partial f(\xi_1, x_2 + y_2 h)}{\partial x_1} y_1 + \frac{\partial f(x_1, \xi_2)}{\partial x_2} y_2$$

Since $x_1 + y_1 h \to x_1$ as $h \to 0$, $\xi_1 \to x_1$ and $\xi_2 \to x_2$. Then taking limits in h yields

$$D_y f(x) = \frac{\partial f(x_1, x_2)}{\partial x_1} y_1 + \frac{\partial f(x_1, x_2)}{\partial x_2} y_2$$

Note that even through y is not an "infinitesimal vector", this is the differential change in the *value* of $f(x)$ with respect to an *infinitesimal* change in x in the direction y. In general,

Theorem 8.4.2 *The change in* $f(x)$ *in the direction* y *is given by the* directional derivative of $f(x)$ in the direction y,

$$D_y f(x) = \sum_{i=1}^{N} \frac{\partial f(x)}{\partial x_i} y_i$$

or

$$D_y f(x) = \nabla f(x) \cdot y$$

This is certainly useful for maximization purposes: You should anticipate that x^* is a local maximum if, for any y, $D_y f(x^*) = 0$. Notice also that the total differential $df(x)$, is just the direction derivative with respect to the unit vector $e = (dx_1, dx_2, ..., dx_N)$, or

$$df(x) = \nabla f(x) \cdot dx$$

The Hessian

The gradient only generalizes the first derivative, however, and our experience from the one-dimensional case tells us that the second derivative is also important. But Exercise 8 from Chapter 4 suggests that the vector of second partial derivatives of $f(x)$ is not sufficient to determine whether a point is a local maximum or minimum. The correct generalization of the second derivative is not another vector, but a matrix.

Definition 8.4.3 *The* Hessian of $f(x)$ *is the matrix*

$$\begin{bmatrix} \frac{\partial^2 f(x)}{\partial x_1^2} & \frac{\partial^2 f(x)}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f(x)}{\partial x_N \partial x_1} \\ \frac{\partial^2 f(x)}{\partial x_2 \partial x_1} & \frac{\partial^2 f(x)}{\partial x_2^2} & & \\ \vdots & & \ddots & \\ \frac{\partial^2 f(x)}{\partial x_N \partial x_1} & & & \frac{\partial^2 f(x)}{\partial x_N^2} \end{bmatrix}$$

Various notations for the Hessian are $\nabla_x^2 f(x)$, $D^2 f(x)$, $\nabla_{xx} = \nabla_x \cdot \nabla_x'$, *and* $H(x)$.

For notational purposes, it is sometimes easier to write the matrix

$$\nabla_{xx} = \begin{bmatrix} \frac{\partial^2}{\partial x_1^2} & \frac{\partial^2}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2}{\partial x_N \partial x_1} \\ \frac{\partial^2}{\partial x_2 \partial x_1} & \frac{\partial^2}{\partial x_2^2} & & \\ \vdots & & \ddots & \\ \frac{\partial^2}{\partial x_N \partial x_1} & & & \frac{\partial^2}{\partial x_N^2} \end{bmatrix}$$

and think of the Hessian as $\nabla_{xx} * f(x) = H(x)$. In particular, $\nabla_{xx} = \nabla_x \cdot \nabla_x'$, so that when you take the "gradient of a gradient" you get a Hessian.

To begin understanding this matrix, consider the quadratic form,

$$f(x_1, x_2) = a_1 x_1 + a_2 x_2 + \frac{1}{2} b_{11} 2 x_1^2 + \frac{1}{2} b_{22} x_2^2 + b_3 x_1 x_2 + c$$

This is the correct generalization of a quadratic function $bx^2 + ax + c$ to two dimensions, allowing interaction between the x_1 and x_2 arguments. It can be written as

$$f(x) = a'x + \frac{1}{2} x' B x + c$$

where $a = [a_1, a_2]'$ and

$$B = \begin{bmatrix} b_{11} & b_3 \\ b_3 & b_{22} \end{bmatrix}$$

By increasing the rows and columns of a and H, this can easily be extended to mappings from $\mathbb{R}^N$ for any N.

The gradient of $f(x_1, x_2)$ is

$$\nabla f(x) = \begin{bmatrix} a_1 + b_{11} x_1 + b_3 x_2 \\ a_2 + b_{22} x_2 + b_3 x_1 \end{bmatrix}$$

But you can see that x_2 appears in $f_{x_1}(x)$, and vice versa, so that changes in x_2 affect the partial derivative of x_1. To summarize this information, we need the extra, off-diagonal terms of the Hessian, or

$$H(x) = \begin{bmatrix} b_{11} & b_3 \\ b_3 & b_{22} \end{bmatrix} = B$$

A basic fact of importance is that the Hessian is symmetric matrix when $f(x)$ is continuous, or or $H' = H$:

Theorem 8.4.4 (Young's Theorem) *If $f(x)$ is a continuous function from $\mathbb{R}^N \to \mathbb{R}$, then*

$$\frac{\partial^2 f(x)}{\partial x_i \partial x_j} = \frac{\partial^2 f(x)}{\partial x_j \partial x_i}$$

and more generally, the order of partial differentiation can be interchanged arbitrarily for all higher-order partial derivatives.

8.5 Taylor Polynomials in $\mathbb{R}^n$

Second-order approximations played a key role in developing second-order sufficient conditions for the one-dimensional case, and we need a generalization of them for the multi-dimensional case. Again, however, the right "generalization" is somewhat subtle. The second-order Taylor polynomials of the one-dimensional case were quadratic functions with an error term,

$$f(x) = f(x_0) + f'(x_0)(x - x_0) + f''(x_0)\frac{(x - x_0)^2}{2} + o(h^3)$$

In the multi-variate case, we start with a quadratic function,

$$f(x) = c + a'x + \frac{1}{2}x'Bx$$

and replace each piece with its generalization of the one-dimensional derivative and second derivative to get a *quadratic approximation*

$$f(x) = f(x_0) + \nabla f(x_0)(x - x_0) + (x - x_0)'\frac{H(x_0)}{2}(x - x_0)$$

and this quadratic approximation becomes a Taylor polynomial when we add the remainder/error term, $o(h^3)$:

$$f(x) = f(x_0) + \nabla f(x_0)(x - x_0) + (x - x_0)'\frac{H(x_0)}{2}(x - x_0) + o(h^3)$$

where $h = x - x_0$.

Definition 8.5.1 *Let $f : D \to \mathbb{R}$ be a thrice-differentiable function, and D an open subset of $\mathbb{R}^N$. Then the* second-order Taylor polynomial of $f(x)$ at x_0 *is*

$$f(x) = f(x_0) + \nabla f(x_0)(x - x_0) + \frac{1}{2}(x - x_0)'H(x_0)(x - x_0) + o(h^3)$$

where $h = x - x_0$, and the remainder $o(h^3) \to 0$ *as $x \to x_0$.*

A *linear approximation* of f around a point x_0 is the function

$$f(x) = f(x_0) + \nabla f(x_0)(x - x_0)$$

and a *quadratic approximation* of f around a point x_0 is the function

$$f(x) = f(x_0) + \nabla f(x_0)(x - x_0) + (x - x_0)'\frac{H(x_0)}{2}(x - x_0)$$

Higher-order approximations require a lot more notation, but second-order approximations are all we need.

8.6 Definite Matrices

In the one dimensional case, the quadratic function

$$f(x) = c + ax + Ax^2$$

was well-behaved for maximization purposes when it had a negative second derivative, or $A < 0$. In the multi-variate case,

$$f(x) = c + a \cdot x + \frac{1}{2} x' Ax$$

it is unclear what sign

$$x' Ax$$

takes. For instance, $a \cdot x > 0$ if $a \geq 0$ and $x \geq 0$. But when is $x' Ax \geq, \leq, >, < 0$? If we knew this, it would be easier to pick out "nice" optimization problems, just as we know that the one-dimensional function $f(x) = Ax^2 + ax + c$ will only have a "nice" solution if $A < 0$.

We can easily pick out some matrices A so that $x' Ax < 0$ for all $x \in \mathbb{R}^N$. If

$$A = \begin{bmatrix} a_1 & 0 & 0 & 0 \\ 0 & a_2 & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & a_N \end{bmatrix}$$

then $x' Ax = \sum_{i=1}^{N} a_i x_i^2$, and this will only be negative if each $a_i < 0$. Notice that,

1. The eigenvalues of A are precisely the diagonal terms $a_1, ..., a_N$
2. The absolute value of a_i is greater than the sum of all the off-diagonal terms (which is zero)
3. If we compute the determinant of each sub-matrix starting from the upper left-hand corner, the determinants will alternate in sign

These are all characterizations of a *negative definite matrix*, or one for which $x' Ax < 0$ for all x. But once we start adding off-diagonal terms, this all becomes much more complicated because it is no longer sufficient merely to have negative terms on the diagonal. Consider the following matrix:

$$A = \begin{bmatrix} -1 & -2 \\ -2 & -1 \end{bmatrix}$$

All the terms are negative, so many students unfamiliar with definite matrices think that $x' Ax$ will then be negative, as an extension of their intuitions about one-dimensional quadratic forms $ax^2 + bx$. But this is false. Take the vector $x = [1, -1]$:

$$x' Ax = [1, -1]' [1, -1] = 1 + 1 = 2 > 0$$

The failure here is that the quadratic form $x' Ax = -x_1^2 + -x_2^2 - 4x_1 x_2$ looks like a "saddle", not like a "hill". If $x_1 x_2 < 0$, then $-4x_1 x_2 > 0$, and this can wipe out $-x_1^2 - x_2^2$, leaving a positive number. So a negative definite matrix must have a negative diagonal, and the "contribution" of the diagonal terms will outweigh those of the off-diagonal terms.

Let A be a $N \times N$ matrix, and $x = (x_1, x_2, ..., x_n)$ a vector in $\mathbb{R}^N$. A *quadratic form* on $\mathbb{R}^N$ is the function

$$x' Ax = \sum_{i=1}^{N} \sum_{j=1}^{N} a_{ij} x_i x_j$$

Definition 8.6.1 *Let A be an $n \times n$ symmetric matrix. Then A is*

- positive definite *if, for any $x \in \mathbb{R}^n \backslash 0$, $x'Ax > 0$*
- negative definite *if, for any $x \in \mathbb{R}^n \backslash 0$, $x'Ax < 0$*
- positive semi-definite *if, for any $x \in \mathbb{R}^n \backslash 0$, $x'Ax \geq 0$*
- negative semi-definite *if, for any $x \in \mathbb{R}^n \backslash 0$, $x'Ax \leq 0$*

Why does this relate to optimization theory? When maximizing a multi-dimensional function, a local (second-order Taylor) approximation of it around a point x^* should then look like a hill if x^* is a local maximum. If, locally around x^*, the function looks like a saddle or a bowl, the value of the function could be increased by moving away from x^*. So these tests are motivated by the connections between the geometry of maximization and quadratic forms. In short, these definitions generalize the idea of a positive or negative number in $\mathbb{R}$ to a "positive or negative matrix".

Definition 8.6.2 *The* leading principal minors *of a matrix A are the matrices,*

$$
\begin{aligned}
A_1 &= a_{11} \\
A_2 &= \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \\
&\vdots \\
A_n &= \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \vdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix}
\end{aligned}
$$

Here are the main tools for characterizing definite matrices:

Theorem 8.6.3 *Let A be an $n \times n$ symmetric matrix. Then A is negative definite if*

- *All the eigenvalues of A are negative*
- *The leading principal minors of A alternate in sign, starting with* $\det(A_1) < 0$
- *The matrix has a* negative dominant diagonal*: For each row r,*

$$|a_{rr}| > \left| \sum_{i \neq r} a_{ri} \right|$$

and $a_{rr} < 0$.

Example Consider

$$A = \begin{bmatrix} -2 & 1 \\ 1 & -2 \end{bmatrix}$$

The leading principal minors are $-2 < 0$ and $4 - 1 = 3 > 0$, so this matrix is negative definite.

The eigenvalues are found by solving

$$\det(A - \lambda I) = (-2 - \lambda)^2 - 1 = 0$$

and $\lambda^* = -2 \pm 1 < 0$. So the eigenvalues of A are all negative.

Since $2 > 1$, it has the dominant diagonal property. ∎

Example Consider

$$A = \begin{bmatrix} -1 & -2 \\ -2 & -1 \end{bmatrix}$$

The leading principal minors are $-1 < 0$ and $1 - 4 < 0$, so this matrix is not negative definite.

The eigenvalues are found by solving

$$\det(A - \lambda I) = \lambda^2 + 2\lambda - 3 = 0$$

and $\lambda^* = -1 \pm 2$. It is not negative definite since it has an eigenvalue $1 > 0$.

The principal minors test is obviously not satisfied.

Since A is neither negative nor positive (semi-)definite, it is actually *indefinite*. ∎

These tests seem "magical", so it's important to develop some intuition about what they mean.

- If H is a real symmetric matrix, it can be factored into an orthogonal matrix P and a diagonal matrix D where
$$H = PDP'$$
and the eigenvalues $(\lambda_1, \lambda_2, ...\lambda_n)$ of H are on the diagonal. If we do this,
$$x'Hx = x'PDP'x = (x'P)D(P'x) = (P'x)'D(P'x)$$
Note that $P'x$ is a vector; call it y. Then we get
$$y'Dy = \sum_{\ell} y_i^2 \lambda_i$$
If all the eigenvalues λ_i of H are negative, the y_i^2 terms are positive scalars, and we can conclude $x'Hx < 0$ if and only if H has all strictly negative eigenvalues.

- How can the principal minors test be motivated? We can get a feel for why this must be true from the following observations: (i) If $x'Hx < 0$ for all x, then if we use $\tilde{x} = (x_1, x_2, ..., x_k, 0, ..., 0)$, with not all components zero, then $\tilde{x}'H\tilde{x} < 0$ as well. But because of the zeros, we're really just considering a $k \times k$ sub-matrix of H, H_k. (ii) The determinant of a matrix A is equal to the product of its eigenvalues. So if we take H_1, the determinant is h_{11}, and that should be negative; $det(H_2) = \lambda_1\lambda_2$, which should be positive if H is negative definite (since the determinant is the product of the eigenvalues, and we know that the eigenvalues of a negative definite matrix are negative) ; $det(H_3) = \lambda_1\lambda_2\lambda_3 < 0$; and so on. So combining these two facts tells us that if H is negative definite, then we should observe this alternative sign pattern in the determinants of the principal minors. (Note that this only shows that if H is negative definite, then the principal minors test holds, not the converse.)

- Necessity of the dominant diagonal theorem is easy to show, but sufficiency is much harder. Suppose H is negative definite. Take the set of vectors $1 = (1, 1, ..., 1)$, and observe the quadratic form
$$1'H1 = \sum_i \sum_j h_{ij} < 0$$
Now, if for each row i, we had $\sum_j h_{ij} < 0$, the above inequality holds, since we then just sum over all i. But this is equivalent to
$$h_{ii} + \sum_{j \neq i} h_{ji} < 0$$

Re-arranging yields

$$h_{ii} < -\sum_{j \neq i} h_{ji}$$

Multiplying by -1 yields

$$-h_{ii} > \sum_{j \neq i} h_{ji}$$

and taking absolute values yields

$$|h_{ii}| > \left| \sum_{j \neq i} h_{ji} \right|$$

which is the dominant diagonal condition.

There are other tests and conditions available, but these are often the most useful in practice. Note that it is harder to identify negative semi-definiteness, because the only test that still applies is that all eigenvalues be weakly negative. In particular, if all the determinants of the principal minors alternate in sign but are sometimes zero, the quadratic form might be a saddle, similar to $f(x) = x^3$ being indefinite at $x = 0$, with $f''(x) = 6x$.

Exercises

1. Show that $\mathbb{R}^N$ is both open and closed. Show that the union and intersection of any collection of closed sets is closed. Show that the union of any collection of open sets is open, and the intersection of a finite number of open sets is open. Show that a countable intersection of open sets can be closed.

2. Determine if the following matrices are positive or negative (semi)-definite:

$$\begin{bmatrix} -1 & 3 \\ 0 & -1 \end{bmatrix}$$

$$\begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix}$$

$$\begin{bmatrix} -1 & 5 \\ 3 & -10 \end{bmatrix}$$

$$\begin{bmatrix} 1 & -2 & 1 \\ -2 & 4 & -2 \\ 1 & -2 & 1 \end{bmatrix}$$

3. Find the gradients and Hessians of the following functions, evaluate the Hessian at the point given, and determine whether it is positive or negative (semi-)definite. (i) $f(x, y) = x^2 + \sqrt{x}$ at $(1, 1)$. (ii) $f(x, y) = \sqrt{xy}$ at $(3, 2)$, and any (x, y). (iii) $f(x, y) = (xy)^2$ at $(7, 11)$ and any (x, y). (iv) $f(x, y, z) = \sqrt{xyz}$ at $(2, 1, 3)$.

4. Give an example of a matrix that has all negative entries on the diagonal but is not negative definite. Give an example of a matrix that has all negative entries but is not negative definite.

5. Prove that a function is continuous iff the inverse image of an open set is open, and the inverse image of a closed set is closed. Prove that if a function is continuous, then the image of a bounded set is bounded. Conclude that a continuous function maps compact sets to compact sets. Use this to provide a second proof of the extreme value theorem.

Proofs

The proofs here are mostly about compactness and the Heine-Borel theorem.

Definition 8.6.4 *A subset K of Euclidean space is* compact *if each sequence in K has a subsequence that converges to a point in K.*

So a set is compact[3] if the Bolzano-Weierstrass theorem holds for every sequence constructed only from points in the set. For an example of a non-compact set, consider (a, b). The sequence $x_n = b - 1/n$ is constructed completely from points in (a, b), but its limit, b, is not in the set, so it does not have the Bolzano-Weierstrass property.

Recall that in our proof of the Weierstrass theorem, we insured a maximum existed by studying convergent sub-sequences. This will be the key to again ensuring existence in the N-dimensional case. The next proof theorem characterizes "closedness" in terms of sequences.

Theorem 8.6.5 *Suppose x_n is a convergent sequence such that x_n is in S for all n. Then x_n converges to a point in the closure of S, $\bar{S}$.*

Proof (Sketch a picture as you go along). Suppose, by way of contradiction, that $\bar{x}$ is not in $\bar{S}$. Then we can draw an open ball around $\bar{x}$ of radius $\varepsilon > 0$ such that no points of $\bar{S}$ are in $B_\varepsilon(\bar{x})$. But since $x_n \to \bar{x}$, we also know that for all $\varepsilon > 0$, for $n \geq H$, $||x_n - \bar{x}|| < \varepsilon$, so that countably many points of $\bar{S}$ are in $B_\varepsilon(\bar{x})$. This is a contradiction. ∎

Since a closed set satisfies $\bar{S} = S$, this implies that every sequence in a closed set — if it converges — converges to a member of the set. However, there are sets like $[0, \infty)$ that are closed — since any sequence that converges in this set converges to a point of the set — but allow plenty of non-convergent sequences, like $x_n = n$. For example, if you were asked to maximize $f(x) = x$ on the set $[b, \infty)$, no maximum exists: $f([b, \infty)) = [b, \infty)$, and the range is unbounded. So for maximization theory, it appears that closedness and boundedness of the image of a set under a function are key properties. In fact, they are equivalent to compactness in $\mathbb{R}^N$.

Theorem 8.6.6 (Heine-Borel) *In Euclidean space, a set is compact iff it is closed and bounded.*

Proof Consider any sequence x_n in a closed and bounded subset K of $\mathbb{R}^N$. We will show it has a convergent subsequence, and consequently that K is compact.

Since K is bounded, there is a N-dimensional "hypercube" H, which satisfies $H = \times_{i=1}^N [a_i, b_i]$ and $K \subset H$. Without loss of generality we can enlarge the hypercube to $[a, b]^N$, where $a = \min_i a_i$ and $b = \max_i b_i$, so that $b_i - a_i$ is the same length for all dimensions.

Now cut H in N^2 equally sized sets, each of size $((b - a)/N)^N$. One of these cubes contains an infinite number of terms of the sequence in K. Select a term from that cube and call it x_{n_1}, and throw the rest of the cubes away. Now repeat this procedure, cutting the remaining cube N ways along each dimensions; one of these sub-cubes contains an infinite number of terms of the sequence; select a term from that subcube and call it x_{n_k}, and throw the rest away.

The volume of the subcubes at each step k is equal to

$$\frac{(b-a)^N}{N^k}$$

[3] The most general definition of compactness is, "A subset K of Euclidean space is *compact* if any collection of open sets $\{O_i\}_{i=1}^\infty$ for which $K \subset \cup_i O_i$ has a finite collection $\{O_{i_k}\}_{k=1}^K$ so that $K \subset \cup_{k=1}^K O_{i_k}$", which is converted into words by saying, "K is compact if every open cover of K — $K \subset \cup_{i=1}^\infty O_i$ — has a finite sub-cover — $K \subset \cup_{k=1}^K O_{i_k}$", or that if an infinite collection of open sets covers K, we can find a finite number of them that do the same job. This is actually easier to work with than the sequential compactness we'll use, but they are equivalent for ^{N}.

which is clearly converging to zero as $k \to \infty$. Then a bound on the distance from each term to all later terms in the sequence is given by the above estimate.

Therefore, the sequence constructed from this procedure has a limit, $\bar{x}$ (since it is a Cauchy sequence). Therefore, the subsequence x_{n_k} converges. Since K is closed, it contains all of its limit points by Theorem 8.8.2 above, so $\bar{x}$ is an element of K. Therefore K is compact. ∎

The Bolzano-Weierstrass theorem was a statement about bounded sequences: Every bounded sequence has a convergent subsequence. The Heine-Borel theorem is a statement about closed and bounded sets: A set is compact iff it is closed and bounded. The bridge between the two is that an infinitely long sequence in a bounded set must be near some point $\bar{x}$ infinitely often. If the set contains all of its points of closure, this point $\bar{x}$ is actually a member of the set K, and compactness and boundedness are closely related. However, boundedness is not sufficient, since a sequence might not converge to a point in the set, like the case (a, b) with $b - 1/n$. To ensure the limit is in the set, we add the closedness condition, and we get a useful characterization of compactness.

Chapter 9

Unconstrained Optimization

We've already seen some examples of unconstrained optimization problems. For example, a firm who faces a price p for its good, hired capital K and labor L at rates r and w per unit, and produces output according to the production technology $F(K, L)$ faces an unconstrained maximization problem

$$\max_{K,L} pF(K,L) - rK - wL$$

In this and similar problems, we usually allow the choice set to be $\mathbb{R}^N$, and allow the firm to pick any K and L that maximizes profits. As long as K and L are both positive, this is a fine approach, and we don't have to move to the more complicated world of constrained optimization.

Similarly, many constrained problems have the feature that the constraint can be re-written in terms of one of the controls, and substituted into the objective. For example, in the consumer's problem

$$\max_{x} u(x)$$

subject to $w = p \cdot x$, the constraint can be re-written as

$$x_N = \frac{w - \sum i = 1^{N-1} p_i x_i}{p_N}$$

and substituted into the objective to yield

$$\max_{x_1,...,x_{N-1}} u\left(x_1, ..., x_{N-1}, \frac{w - \sum i = 1^{N-1} p_i x_i}{p_N}\right)$$

which is an unconstrained maximization problem in $x_1, ..., x_{N-1}$. Even though we have better methods in general for solving the above problem from a theoretical perspective, solving an unconstrained problem numerically is generally easier that solving a constrained one.

Definition 9.0.7 *Let $f : \mathbb{R}^n \to \mathbb{R}$, $x \in \mathbb{R}^n$. Then the* unconstrained maximization problem *is*

$$\max_{x \in \mathbb{R}^N} f(x)$$

A global maximum of f on X *is a point x^* so that $f(x^*) \geq f(x')$ for all other $x' \in X$. A point x^* is a* local maximum of f *if there is an open set $S = \{y : ||x - y|| < \varepsilon\}$ for which $f(x^*) \geq f(y)$ for all $y \in S$.*

How do we find solutions to this problem?

9.1 First-Order Necessary Conditions

Our first step to solving unconstrained maximization problems is to build up a candidate list using FONCs, just like in the one-dimensional case.

Definition 9.1.1 *If $\nabla f(x^*) = 0$, then x^* is a* critical point *of $f(x)$.*

Example Recall the firm with profit function

$$\pi(K, L) = \max_{K,L} pF(K, L) - rK - wL$$

Then if we make a small change in K, the change in profits is

$$\frac{\partial \pi(K, L)}{\partial K} = pF_K(K, L) - r$$

and if we make a small change in L, the change in profits is

$$\frac{\partial \pi(K, L)}{\partial L} = pF_L(K, L) - w$$

If there are no profitable adjustments away from a given point (K^*, L^*), then it must be a local maximum, so the equations above both equation zero. But then $\nabla \pi(K^*, L^*) = 0$ implies that (K^*, L^*) is a critical point of $f(x)$. ∎

The above argument works for any function $f(x)$:

Theorem 9.1.2 (First-Order Necessary Conditions) *If x^* is a local maximum of f and f is differentiable at x^*, then x^* is a critical point of f.*

Proof If x^* is a local maximum and f is differentiable at x^*, there cannot be any improvement in any direction $v \neq 0$. The directional derivative is

$$\nabla f(x^*) \cdot v = \sum_i \frac{\partial f(x^*)}{\partial x_i} v_i$$

So we can think of the differential change as the sum of one-dimensional directional derivatives in the direction $y = (0, ..., v_i, 0, ..., 0)$ where y is zero except for the i-th slot, taking the value $v_i \neq 0$:

$$\nabla f(x^*) \cdot y = \sum_i \frac{\partial f(x^*)}{\partial x_i} v_i = \frac{\partial f(x^*)}{\partial x_i} v_i = 0$$

So that each partial derivative must be zero along each dimension individually, implying that a local maximum is a critical point. ∎

Like in the one-dimensional case, this gives us a way of building a candidate list of potential maximizers: Critical points and any points of non-differentiability.

Example Consider a quadratic objective function,

$$f(x_1, x_1) = a_1 x_1 + a_2 x_2 - \frac{b_1}{2} x_1^2 - \frac{b_2}{2} x_2^2 + c x_1 x_2$$

The FONCs are

$$a_1 - b_1 x_1^* + c x_2^* = 0$$

$$a_2 - b_2 x_2^* + c x_1^* = 0$$

Any solution to these equations is a critical point. If we solve the system by hand, the second equation is equivalent to

$$x_2^* = \frac{a_2 + c x_1^*}{b_2}$$

and substituting it into the first gives

$$a_1 - b_1 x_1^* + c\frac{a_2 + c x_1^*}{b_2} = 0$$

or

$$x_1^* = \frac{a_1 b_2 + a_2 c}{b_1 b_2 - c^2}$$

$$x_2^* = \frac{a_2 b_1 + a_1 c}{b_1 b_2 - c^2}$$

If, instead, we convert this to a matrix equation,

$$\begin{bmatrix} a_1 \\ a_2 \end{bmatrix} + \begin{bmatrix} -b_1 & c \\ c & -b_2 \end{bmatrix} \begin{bmatrix} x_1^* \\ x_2^* \end{bmatrix} = 0$$

Re-arranging yields

$$-\begin{bmatrix} -b_1 & c \\ c & -b_2 \end{bmatrix} \begin{bmatrix} x_1^* \\ x_2^* \end{bmatrix} = \begin{bmatrix} a_1 \\ a_2 \end{bmatrix}$$

Which looks exactly like $-Bx^* = a$. From linear algebra, we know that there is a solution as long as B is non-singular (it has full rank $\leftrightarrow$ all its eigenvalues are non-zero $\leftrightarrow$ it is invertible $\leftrightarrow$ it has non-zero determinant). Then

$$x^* = (-B)^{-1} a$$

and we have a solution. The determinant of B is non-zero iff

$$b_1 b_2 - c^2 \neq 0$$

So if the above condition is satisfied, there is a unique critical point.

Notice, however, that we could have done the whole analysis as

$$f(x) = ax + x' \frac{B}{2} x$$

yielding FONCs

$$a + Bx^* = 0$$

or

$$x^* = -B^{-1} a$$

Moreover, this equation can be solved in a few lines of Matlab code, making it useful starting point for playing with non-linear optimization problems. However, we still don't know know whether x^* is a maximum, minimum, or saddle point. $\blacksquare$

9.2 Second-Order Sufficient Conditions

With FONCs, we can put together a candidate list for any unconstrained maximization problem: Critical points and any points of non-differentiability. However, we still don't know whether a given critical point is maximum, minimum or saddle/inflection point.

Example Consider the function

$$f(x, y) = -x^2 y$$

The FONCs for this function are

$$f_x(x, y) = -2xy = 0$$

$$f_y(x, y) = -x^2 = 0$$

The unique critical point is $(0, 0)$. But if we evaluate the function at $(1, -1)$, we get $f(1, -1) = -(1)(-1) = 1 > 0 = f(0, 0)$. What is going on? Well, if we were considering $-x^2$ alone, it would have a global maximum at zero. But the y term has no global maximum. When you multiply these functions together, even though one is well-behaved, the mis-behavior of the other leads to non-existence of a maximizer in $\mathbb{R}^2$ (I can make $f(x, y)$ arbitrarily large by making y arbitrarily negative and $x = 1$). ∎

Still worse, as you showed in Exercise 8 of Chapter 4, it is not enough that a critical point (x^*, y^*) satisfy $f_{xx}(x^*, y^*) < 0$ and $f_{yy}(x^*, y^*) < 0$ to be a local maximum.

Example Consider the function

$$f(x, y) = -\frac{1}{2}x^2 - \frac{1}{2}y^2 + bxy$$

where $b > 0$. The FONCs are

$$f_x(x, y) = -x + by = 0$$

$$f_y(x, y) = -y + bx = 0$$

The unique critical point of the system is $(0, 0)$. However, if we set $x = y = z$, the function becomes

$$f(z) = -z^2 + bz^2$$

and if $b > -1$, we can make the function arbitrarily large. The function is perfectly well behaved in each direction: If you plot a cross-section of the function for all x setting y to zero, it achieves a maximum in x at zero, and if you plot a cross-section of the function for all y setting x to zero, it achieves a maximum in y at zero. Nevertheless, this is not a global maximum if b is too large. ∎

If we use a Taylor series, however, we can write any multi-dimensional function as

$$f(x) = f(x_0) + \nabla f(x_0) \cdot (x - x_0) + (x - x_0)' \frac{H(x_0)}{2}(x - x_0) + o(h^3)$$

If x^* is a critical point, we know that $\nabla f(x^*) = 0$, or

$$f(x) = f(x^*) + (x - x^*) \cdot \frac{H(x^*)}{2}(x - x^*) + o(h^3)$$

and re-arranging yields

$$f(x^*) - f(x) = -(x - x^*) \cdot \frac{H(x^*)}{2}(x - x^*) - o(h^3)$$

Letting $h = x - x^*$ be arbitrarily close to zero and noting that if x^* is a local maximum of $f(x)$, we then have

$$f(x^*) - f(x) = -(x - x^*) \cdot \frac{H(x^*)}{2}(x - x^*) > 0$$

Or that, for any vector y,

$$y' H(x^*) y < 0$$

This is the definition of a negative definite matrix, giving us a workable set of SOSCs for multi-dimensional maximization:

Theorem 9.2.1 *(Second-Order Sufficient Conditions) If x^* is a critical point of $f(x)$ and $H(x^*)$ is negative definite, then x^* is a local maximum of $f(x)$.*

For example, the Hessian for

$$f(x, y) = -\frac{1}{2}x^2 - \frac{1}{2}y^2 + bxy$$

is

$$\begin{bmatrix} -1 & b \\ b & -1 \end{bmatrix}$$

The determinant of the Hessian is $1 - b^2$, which is positive if $b < 1$. So if b is sufficiently small, $f(x, y)$ will have a local maximum at $(0, 0)$.

On the other hand, if we start with the assumption that x^* is a local maximum of $f(x)$ and not just a critical point, using our Taylor series approximation, we get

$$f(x^*) - f(x) = -(x - x^*)' \frac{H(x^*)}{2}(x - x^*) + o(h^3)$$

Since x^* is a local maximum, we know that $f(x^*) \geq f(x)$, so that

$$-(x - x^*)' \frac{H(x^*)}{2}(x - x^*) \geq 0$$

for x sufficiently close to x^*, implying that $y' H(x^*) y \leq 0$. This is negative *semi*-definiteness.

To summarize:

Theorem 9.2.2 (Second-Order Sufficient Conditions) *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a twice differentiable function, and let x^* be a point in $\mathbb{R}^n$.*

- *If x^* is a critical point of $f(x)$ and $H(x^*)$ is negative definite, then x^* is a local maximum of f.*
- *If x^* is a critical point of $f(x)$ and $H(x^*)$ is positive definite, then x^* is a local minimum of f.*
- *If f has a local maximum at x^*, then $H(x^*)$ is negative semi-definite.*
- *If f has a local minimum at x^*, then $H(x^*)$ is positive semi-definite.*

The first two points are useful for checking whether or not a particular point is a local maximum or minimum. The second two are useful when using the implicit function theorem.

Here are some examples:

Example Suppose a price-taking firm gets a price p for its product, which it produces using capital K and labor L and technology $F(K, L) = q$. Capital costs r per unit and labor costs w per unit. This gives a profit function of

$$\pi(K, L) = pF(K, L) - rK - wL$$

The FONCs are

$$pF_K(K, L) - r = 0$$

$$pF_L(K, L) - w = 0$$

And our SOSCs are that

$$\begin{bmatrix} F_{KK}(K^*, L^*) & F_{KL}(K^*, L^*) \\ F_{LK}(K^*, L^*) & F_{LL}(K^*, L^*) \end{bmatrix}$$

be a negative-definite matrix. This implies

$$F_{KK}(K^*, L^*) < 0$$

$$F_{LL}(K^*, L^*) < 0$$

and

$$F_{KK}(K^*, L^*)F_{LL}(K^*, L^*) - F_{KL}(K^*, L^*)^2 > 0$$

So $F(K, L)$ must be *own-concave* in K and L, or $F_{KK}(K^*, L^*) < 0$ and $F_{LL}(K^*, L^*) < 0$. Likewise, the cross-partial $F_{KL}(K^*, L^*)$ cannot be "too large". In terms of economics, K and L cannot be such strong substitutes or complements that switching from one to the other has a larger impact than using more of each. A simple example might be

$$F(K, L) = a_1 K - \frac{a_2}{2}K^2 + b_1 L - \frac{b_2}{2}L^2 + cKL$$

Our Hessian would then be

$$\begin{bmatrix} -a_2 & c \\ c & -b_2 \end{bmatrix}$$

with determinant

$$a_2 b_2 - c^2 > 0$$

So that (K^*, L^*) is a local maximum if $\sqrt{a_2 b_2} > c$. ∎

Example Slightly different firm problem: A firm hires capital K at rental rate r and labor L at wage rate w. The firm's output produced by a given mix of capital and labor is $F(K, L) = \phi K^\alpha L^\beta$. The price of the firm's good is p. This is an unconstrained maximization problem in (K, L), so we can solve

$$\max_{K,L} p\phi K^\alpha L^\beta - rK - wL$$

This gives first-order conditions

$$r = p\phi\alpha K^{\alpha-1}L^\beta$$
$$w = p\phi\beta K^\alpha L^{\beta-1}$$

Dividing these two equations yields

$$\frac{r}{w} = \frac{\alpha L}{\beta K}$$

Solving for K in terms of L yields

$$L = \frac{r}{w}\frac{\beta}{\alpha}K$$

Substituting back into the first-order condition for K gives

$$r = p\phi\alpha K^{\alpha-1}\left(\frac{r}{w}\frac{\beta}{\alpha}K\right)^{\beta}$$

So that

$$K^* = \left(p\phi\frac{\alpha^{1+\beta}}{\beta^{\beta}}\frac{w^{\beta}}{r^{1+\beta}}\right)^{\frac{1}{1-\alpha-\beta}}$$

Hypothetically, we could now differentiate K^* with respect to α, β, r, ϕ, or w, to see how the firm's choice of capital varies with economic conditions. But that looks inconvenient, especially for α or β, which appear everywhere and in exponents.

What are the SOSCs? The Hessian is

$$\begin{bmatrix} \alpha(\alpha-1)K^{\alpha-2}L^{\beta} & \alpha\beta K^{\alpha-}L^{\beta-1} \\ \alpha\beta K^{\alpha-}L^{\beta-1} & \beta(\beta-1)K^{\alpha}L^{\beta-2} \end{bmatrix}$$

The upper and lower corners are negative if $0 < \alpha < 1$ and $0 < \beta < 1$. The determinant is

$$\det H = \alpha(\alpha-1)\beta(\beta-1)K^{2\alpha-2}L^{2\beta-2} - \alpha^2\beta^2K^{2\alpha-2}L^{2\beta-2}$$

or

$$\det H = \{(\alpha-1)(\beta-1) - \alpha\beta\}\,\alpha\beta K^{2\alpha-2}L^{2\beta-2}$$

Which is positive if $\alpha\beta - \alpha - \beta + 1 - \alpha\beta > 0$, or $1 > \alpha + \beta$.

So as long as $0 < \alpha < 1$, $0 < \beta < 1$ and $1 > \alpha + \beta$, the Hessian is negative definite, and the critical point (K^*, L^*) is a local maximum. Since it is the only point on the candidate list, it is a global maximum. $\blacksquare$

Example Suppose we have a consumer with utility function $u(q_1, q_2, m) = v(q_1, q_2) + m$ over two goods and money, with wealth constraint $w = p_1q_1 + p_2q_2 + m$. Substituting the constraint into the objective, we get

$$\max_{q_1,q_2} v(q_1, q_2) + w - p_1q_1 - p_2q_2$$

The FONCs are

$$v_1(q_1^*, q_2^*) - p_1 = 0$$
$$v_2(q_1^*, q_2^*) - p_2 = 0$$

and the SOSCs are that

$$\begin{bmatrix} v_{11}(q_1^*, q_2^*) & v_{21}(q_1^*, q_2^*) \\ v_{12}(q_1^*, q_2^*) & v_{22}(q_1^*, q_2^*) \end{bmatrix}$$

be negative definite. Can we figure out how purchases of q_1^* respond to a change in p_2? Well, the two functions $q_1^*(p_2)$ and $q_2^*(p_2)$ are implicitly determined by the system of FONCs. If we totally differentiate, we get

$$v_{11}\frac{\partial q_1}{\partial p_2} + v_{21}\frac{\partial q_2}{\partial p_2} = 0$$

$$v_{12}\frac{\partial q_1}{\partial p_2} + v_{22}\frac{\partial q_2}{\partial p_2} - 1 = 0$$

Solving the first equation in terms of $\partial q_2/\partial p_2$, we get

$$\frac{\partial q_2}{\partial p_2} = -\frac{v_{11}\dfrac{\partial q_1}{\partial p_2}}{v_{21}}$$

and substituting into the second equation gives

$$v_{12}\frac{\partial q_1}{\partial p_2} - v_{22}\frac{v_{11}\dfrac{\partial q_1}{\partial p_2}}{v_{21}} - 1 = 0$$

and solving for $\partial q_1/\partial p_2$ yields

$$\frac{\partial q_1}{\partial p_2} = \frac{-v_{21}}{v_{11}v_{22} - v_{12}v_{21}}$$

Notice that the denominator is the determinant of $H(q^*)$, so it must be positive. Consequently, the sign is determined by the numerator,

$$\text{sign}\ \left(\frac{\partial q_1}{\partial p_2}\right) = \text{sign}\ \ (-v_{21})$$

So q_1 are *gross complements* when $v_{21} > 0$, and *gross substitutes* when $v_{21} < 0$. $\blacksquare$

So it is pretty straightforward to apply the implicit function theorem to this two-dimensional problem. But when the number of controls becomes large, it is less obvious that this will work. We will need to develop more subtle tools for working through higher-dimensional comparative statics problems.

9.3 Comparative Statics

Perhaps if we take a broader perspective, we can see some of the structure behind the comparative statics exercise we did for the consumer above. The unconstrained maximization problem

$$\max_x f(x, c)$$

has FONCs

$$\nabla_x f(x^*(c), c) = 0$$

where c is a single exogenous parameter (the extension to a vector of parameters is easy, and in any math econ text).

If we differentiate the FONCs with respect to c, we get

$$\nabla_{xx} f(x^*(c), c) \cdot \nabla_c x^*(c) + \nabla_x f_c(x^*(c), c) = 0$$

Since $\nabla_{xx} f(x^*(c), c) = H(x^*(c), c)$, we can write

$$H(x^*(c), c)\nabla_c x^*(c) = -\nabla_x f_c(x^*(c), c)$$

Since $H(x^*(c), c)$ is negative definite, all its eigenvalues are negative, so it is invertible, and

$$\nabla_c x^*(c) = -H(x^*(c), c)^{-1}(\nabla_x f_c(x^*(c), c))$$

If x and c are both one-dimensional, this is just

$$\frac{\partial x^*(c)}{\partial c} = \frac{f_{xc}(x^*(c), c)}{f_{xx}(x^*(c), c)}$$

So the $H(x^*(c), c)^{-1}$ term is just the generalization of $1/f_{xx}(x^*(c), c)$, and $\nabla_x f_c(x^*(c), c)$ is just the generalization of $f_{xc}(x^*(c), c)$.

Theorem 9.3.1 (Implicit Function Theorem) *Suppose that $\nabla f(x(c), c) = 0$ and that $\nabla f(x, c)$ is differentiable in x and c. Then there is a locally continuous, implicit solution $x^*(c)$ with derivative*

$$\nabla_c x^*(c) = -H(x^*(c), c)^{-1}(\nabla_x f_c(x^*(c), c))$$

Example Recall the profit-maximizing firm with general production function $F(K, L)$, and let's see how a change in r affects K^* and L^*. The system of FONCs is

$$pF_K(K^*, L^*) - r = 0$$

$$pF_L(K^*, L^*) - w = 0$$

Totally differentiating with respect to r yields

$$pF_{KK}\frac{\partial K^*}{\partial r} + pF_{LK}\frac{\partial L^*}{\partial r} - 1 = 0$$

$$pF_{KL}\frac{\partial K^*}{\partial r} + pF_{LL}\frac{\partial L^*}{\partial r} = 0$$

Doing the system "by hand" yields

$$\frac{\partial L^*}{\partial r} = \frac{-pF_{KL}}{F_{KK}F_{LL} - F_{KL}F_{LK}}$$

Note that the denominator is the determinant of the Hessian. ∎

You can always grind the solution out by hand, and I like to do it sometimes to check my answer. But there is another tool, *Cramer's rule*, for solving equations like this. Writing the system of equations in matrix notation gives

$$\underbrace{\begin{bmatrix} pF_{KK} & pF_{KL} \\ pF_{LK} & pF_{LL} \end{bmatrix}}_{\nabla_{xx} f(x^*(c),c)=H(x^*(c))} \underbrace{\begin{bmatrix} \partial K^*/\partial r \\ \partial L^*/\partial r \end{bmatrix}}_{\nabla_c x^*(c)} = \underbrace{\begin{bmatrix} 1 \\ 0 \end{bmatrix}}_{\nabla_x f_c(x^*(c),c)}$$

So we have a matrix equation, $Ax = b$ and we want to solve for x, the vector of comparative statics.

Theorem 9.3.2 (Cramer's Rule) *Consider the matrix equation*

$$Ax = (A_{.1}, A_{.2}, ..., A_{.N}) \begin{bmatrix} x_1 \\ \vdots \\ x_{i-1} \\ x_i \\ x_{i+1} \\ \vdots \\ x_N \end{bmatrix} = b$$

where $A_{.k}$ is the k-th column of A. Then

$$x_i = \frac{\det([A_{.1}, ..., A_{.i-1}, b, A_{.i+1}, ..., A_{.n}])}{\det(A)}$$

So to use Cramer's rule to solve for the i-th component of x, replace the i-th column of A with b and compute that determinant, then divide by the determinant of A.

For the firm example, we get

$$\frac{\partial L^*}{\partial r} = \frac{\det \begin{pmatrix} pF_{KK} & 1 \\ pF_{KL} & 0 \end{pmatrix}}{\det(H)} = \frac{-pF_{KL}}{\det(H)}$$

So the sign of $\partial L^*/\partial r$ is equal to the sign of F_{KL}.

In fact, this should always work, because

$$\underbrace{H(x^*(c), c)}_{N\times N \text{ matrix}} \underbrace{\nabla_c x^*(c)}_{N\times 1 \text{ vector}} = \underbrace{\nabla_x f_c(x^*(c), c)}_{N\times 1 \text{ vector}}$$

can always be written as a matrix equation $Hx = b$. Since H is symmetric, it has all real eigenvalues, so should be invertible.

9.4 The Envelope Theorem

Similarly, we can characterize an envelope theorem for unconstrained maximization problems that shows how an agent's payoff varies with an exogenous parameter. The generic unconstrained maximization problem is

$$\max_x f(x, c)$$

with FONCs

$$\nabla_x f(x^*(c), c) = 0$$

If we consider the *value function*,

$$V(c) = f(x^*(c), c)$$

this gives the maximized payoff of the agent for each value of the exogeneous parameter c. Differentiating with respect to c yields

$$\nabla_c V(c) = \underbrace{\nabla_x f(x^*(c), c)}_{\text{FONC}} \cdot \nabla_c x^*(c) + \nabla_c f(x^*(c), c)$$

Again, since the FONC must be zero at a maximum, we have

$$\nabla_c V(c) = \nabla_c f(x^*(c), c)$$

So the derivative of the value function with respect to a given parameter is just the partial derivative with respect to that parameter.

Example Consider a consumer with utility function

$$V(p_1, p_2, w) = v(q_1^*, q_2^*) + w - p_1 q_1^* - p_2 q_2^*$$

If we differentiate with respect to, say, p_2, each maximized value q_1^* and q_2^* must also be differentiated, yielding

$$\frac{\partial V}{\partial p_2} = v_1 \frac{\partial q_1^*}{\partial p_2} + v_2 \frac{\partial q_2^*}{\partial p_2} - p_1 \frac{\partial q_1^*}{\partial p_2} - p_2 \frac{\partial q_2^*}{\partial p_2} - q_2^*$$

Which is a mess. It would appear that we have to go back and compute all the comparative statics to sign this (and even then there would be no guarantee that would work). But re-arranging yields

$$\frac{\partial V}{\partial p_2} = (v_1 - p_1) \frac{\partial q_1^*}{\partial p_2} + (v_2 - p_2) \frac{\partial q_2^*}{\partial p_2} - q_2^*$$

Using the FONCs, we get

$$\frac{\partial V}{\partial p_2} = -q_2^*$$

This is the logic of the envelope theorem. So we can skip all the intermediate re-arranging steps, and just differentiate directly with respect to parameters once the payoff-maximizing behavior has been substituted in:

$$\frac{\partial V}{\partial w} = 1$$
$$\frac{\partial V}{\partial p_1} = -q_1^*$$

$\blacksquare$

Exercises

1. Prove that if x^* is a local maximizer of $f(x)$, then it is a local minimizer of $-f(x)$.

2. Suppose x^* is a global maximizer of $f : \mathbb{R}^n \to \mathbb{R}$. Show that for any monotone increasing transformation $g : \mathbb{R} \to \mathbb{R}$, x^* is a maximizer of $g(f(x))$.

3. Find all critical points of $f(x, y) = (x^2 - 4)^2 + y^2$ and show which are maxima and which are minima.

4. Find all critical points of $f(x, y) = (y - x^2)^2 - x^2$ and show which are maxima and which are minima.

5. Describe the set of maximizers and set of minimizers of

$$f(x, y) = cos\left(\sqrt{x^2 + y^2}\right)$$

6. Suppose a firm produces two goods, q_1 and q_2, whose prices are p_1 and p_2, respectively. The costs of production are $C(q_1, q_2)$. Characterize profit-maximizing output and show how q_1 varies with p_2. If $C(q_1, q_2) = c_1(q_1) + c_2(q_2) + bq_1q_2$, explain when a critical point is a local maximum of the profit function. How do profits vary with b and p_1?

7. Suppose you have a set of dependent variables $\{y_1, y_2, ..., y_N\}$ generated by independent variables $\{x_1, x_2, ..., x_N\}$, and believe the true model is given by

$$y_i = \beta' x_i + \varepsilon_i$$

where ε is a normally distributed random variable with mean m and variance σ^2. The sum of squared errors is

$$(y - (\beta' x))'(y - (\beta' x)) = \sum_{i=1}^{N} (y_i - \beta' x_i)^2$$

Check that this is a quadratic programming problem. Compute the gradient and Hessian. Solve for the optimal estimator $\hat{\beta}$. (This is just OLS, right?)

8. A consumer with utility function $u(q_1, q_2, m) = (q_1 - \gamma_1)q_2^{\alpha} + m$ and budget constraint $w = p_1q_2 + p_2q_2 + m$ is trying to maximize utility. Solve for the optimal bundle (q_1^*, q_2^*, m^*) and show how q_1^* varies with p_2, and how q_2^* varies with p_1, both using the closed-form solutions and the IFT. How does the value function vary with γ_1?

9. Suppose you are trying to maximize $f(x_1, x_2, ..., x_N)$ subject to the non-linear constraint that $g(x_1, ..., x_N) = 0$. Use the Implicit Function Theorem to (i) use the constraint to define $x_N(x_1, ..., x_{N-1})$ and substitute this into f to get an unconstrained maximization problem in $x_1, ..., x_{N-1}$, (ii) derive FONCs for $x_1, ..., x_{N-1}$.

Chapter 10

Equality Constrained Optimization Problems

Optimization problems often come with extra conditions that the solution must satisfy. For example, consumers can't spend more than their budget allows, and firms are constrained by their technology. Some examples are:

- The canonical equality constrained maximization problem comes from consumer theory. There is a consumer choosing between bundles of x_1 and x_2. He has a *utility function* $u(x_1, x_2)$, which is increasing and differentiable in both arguments. However, he only has wealth w and the prices of x_1 and x_2 per unit are p_1 and p_2, respectively, so that his *budget constraint* is $w = p_1x_1 + p_2x_2$. Then his maximization problem is

$$\max_{x_1,x_2} u(x_1, x_2)$$

  subject to $w = p_1x_1+p_2x_2$. Here, the objective function is non-linear in x, while the constraint is linear in x.

- Consider a firm who transforms inputs z into output q through a technology $F(z) = q$. The cost of input z_k is w_k. The firm would like to minimize costs subject to producing a certain amount of output, $\bar{q}$. Then his problem becomes

$$\min_{z} w \cdot z$$

  subject to $F(z) = \bar{q}$. This is a different problem from the consumer's, primarily because since the objective is linear in z while the constraint is non-linear in z.

These are *equality constrained maximization problems* because the set of feasible points is described by an equation, $w = p_1x_1 + p_2x_2$ (as opposed to an inequality constraint, like $w \geq p_1x_1 + p_2x_2$).

To provide a general theory for all the constrained maximization problems we might encounter, then, we need to write the constraints in a common form. Often, the constraints will have the form $f(x) = c$, where x are the choice variables and c is a constant.

- For maximization problems, move all the terms to the side with the choice variables, and define a new function

$$g(x) = f(x) - c$$

  Then whenever $g(x) = 0$, the constraints are satisfied.

- For minimization problems, move all the firms to the side with the constant, and define a new function

$$g(x) = c - f(x)$$

Then whenever $g(x) = 0$, the constraints are satisfied.

This will ensure that the Lagrange multiplier — see below — is always positive, so that you don't have to figure out its sign later on. For example, the constraint $w = p_1 x_1 + p_2 x_2$ becomes $0 = p_1 x_1 + p_2 x_2 - w = g(x)$.

Definition 10.0.1 *Let $f : D \to \mathbb{R}$. Suppose that the choice of x is subject to an* equality constraint*, such that any solution must satisfy $g(x) = 0$ where $g : D \to \mathbb{R}$. Then the* equality-constrained maximization problem *is*

$$\max_x f(x)$$

subject to

$$g(x) = 0$$

Simply differentiating $f(x)$ with respect to x is no longer a sensible approach to finding a maximizer, since that ignores the constraints. We need a theory that incorporates constraints into the search for a maximum.

10.1 Two useful but sub-optimal approaches

There are two ways of approaching the question of constrained maximization that are instructive, but not necessarily efficient. If you look at proof of Lagrange's theorem, these ideas show up, however, and they give a lot of intuition about how this kind of maximization problem works.

10.1.1 Using the implicit function theorem

Note that the constraint

$$g(x_1, x_2, ..., x_{N-1}, x_N) = 0$$

can be used to formulate an implicit function, $x_N(x_1, x_2, ..., x_{N-1})$, defined by

$$g(x_1, x_2, ..., x_{N-1}, x_N(x_1, x_2, ..., x_{N-1})) = 0$$

Then the unconstrained problem in $x_1, ..., x_{N-1}$ can be stated as:

$$\max_{x_1,...,x_{N-1}} f(x_1, x_2, ..., x_{N-1}, x_N(x_1, x_2, ..., x_{N-1}))$$

with first-order necessary conditions for $k = 1, ..., N-1$,

$$\frac{\partial f(x^*)}{\partial x_k} + \frac{\partial f(x^*)}{\partial x_N} \frac{\partial x_N(x^*)}{\partial x_k} = 0$$

Using the implicit function theorem on the constraint, we get

$$\frac{\partial g(x^*)}{\partial x_k} + \frac{\partial g(x^*)}{\partial x_N} \frac{\partial x_N(x^*)}{\partial x_k} = 0$$

Substituting this into the FONC for k, we get

$$\frac{\partial f(x^*)}{\partial x_k} - \frac{\partial f(x^*)}{\partial x_N} \frac{\partial g(x^*)/\partial x_k}{\partial g(x^*)/\partial x_N} = 0$$

Yielding the tangency conditions

$$\frac{\partial f(x^*)/\partial x_k}{\partial f(x^*)/\partial x_N} = \frac{\partial g(x^*)/\partial x_k}{\partial g(x^*)/\partial x_N} = 0$$

which is a generalization of the familiar "marginal utility of x over marginal utility of y equals the price ratio" relationship in consumer theory.

However, developing and verifying second-order sufficient conditions using this approach appears to be quite challenging. We would need to apply the implicit function theorem to the system of FONC's, leading to a very complicated Hessian that might not obviously be negative definite.

10.1.2 A more geometric approach

Since we are using calculus, let's focus on what must be true locally for a point x^* to maximize $f(x)$ subject to $g(x) = 0$. In particular, imagine that instead of maximizing $f(x)$ in $\mathbb{R}^N$ but being restricted to the points such that $g(x) = 0$, imagine that the set of points such that $g(x) = 0$ is the set over which we maximize $f(x)$.

What do I mean by that? Let $g(x_0) = 0$. The set of *feasible local variations on* x_0 are the set of points for which the directional derivative of $g(x)$ at x_0 evaluated in the direction x' are zero:

$$\nabla g(x_0) \cdot x' = 0$$

This is the set of points x' for which the constraint is still satisfied if a differential step is taken in that direction: Imagine standing at the point x_0 and taking a small step towards x' so that your foot is still in the set of points satisfying $g(x) = 0$. Define this set as

$$Y(x_0) = \{x' : \nabla g(x_0) \cdot x' = 0\}$$

Now, a point x^* is a local maximum of $f(x)$ subject to $g(x) = 0$ if it is a local maximum of $f(x)$ on the set $Y(x^*)$ (right?). This implies the following:

$$\nabla f(x^*) \cdot x' = 0$$

for all x' such that $\nabla g(x^*) \cdot x' = 0$.

This gives great geometric intuition: For any feasible local variation x' on x^*, it must be the case that the gradient of the objective function $f(x)$ and the gradient of the constraint $g(x)$ are both orthogonal to x'.

This supplies more geometric intuition and clarifies the set of points which must consider as potential improvements on a local maximizer (the set of feasible local variations), but doesn't seem to provide an algorithm for solving for maximizers or testing whether they are local maximizers, minimizers, or neither.

10.2 First-Order Necessary Conditions

The preferred method of solving these problems is *Lagrange maximization*. To use this approach, we introduce a special function:

Definition 10.2.1 *The* Lagrangean *is the function*

$$\mathcal{L}(x, \lambda) = f(x) - \lambda g(x)$$

where λ is a real number.

The Lagrangean is designed so that if $g(x) \neq 0$, the term $-\lambda g(x)$ acts as a penalty on the objective function $f(x)$. You might even imagine coming up with an algorithm that works by somehow penalizing the decision-maker for violating constraints by raising the penalties to push them towards a "good" solution that makes $f(x)$ large and satisfies the constraint.

I find that it is best to think of $\mathcal{L}(x, \lambda)$ as a convenient way of converting a constrained maximization problem in x subject to $g(x) = 0$ into an unconstrained maximization problem in terms of (x, λ). When you subtract $\lambda g(x)$ from the objective in the Lagrangian, you are basically imposing an extra cost on the decision-maker for violating the constraint. When you maximizer with respect to λ, you are minimizing the pain of this cost, $\lambda g(x)$ (since maximizing the negative is the same as minimizing). So Lagrange maximization trades off between increasing the objective function, $f(x)$, and satisfying the constraints, $g(x)$, by introducing this fictional cost of violating them.

This is the basic idea of our new first-order necessary conditions:

Theorem 10.2.2 (Lagrange First-Order Necessary Conditions) *If*

1. *x^* is a local maximum of $f(x)$ that satisfies the constraints $g(x) = 0$*
2. *the constraint gradients $g'(x^*) \neq 0$ (this is called the* constraint qualification*)*
3. *$f(x)$ and $g(x)$ are both differentiable at x^**

Then there exists a unique vector λ^* such that the FONCs

$$\nabla_x \mathcal{L}(x^*, \lambda^*) = \nabla f(x^*) - \lambda^* \nabla g(x^*) = 0$$

$$\nabla_\lambda \mathcal{L}(x^*, \lambda^*) = -g(x^*) = 0$$

hold.

Proof Consider $\mathcal{L}(x, \lambda)$ as an unconstrained maximization problem in (x, λ). The first-order necessary conditions are

$$\nabla_x \mathcal{L}(x, \lambda) = \nabla_x f(x) - \lambda' \nabla g(x) = 0$$

$$\nabla_\lambda \mathcal{L}(x, \lambda) = -g(x) = 0$$

Note that the second equation, $-g(x) = 0$, implies that at any critical point of $\mathcal{L}(x, \lambda)$, the constraints $g(x) = 0$ are satisfied. By the implicit function theorem (x^* below is an implicit solution), a local solution to the system of FONCs is

$$\lambda^* = [\nabla g(x^*)']^{-1} \nabla_x f(x^*)$$

$$g(x^*) = 0$$

where $\nabla g(x)$ is the matrix of partial derivatives; this only exists if each row is independent so that the matrix has full rank, so none of the gradients of the constraint vectors can be scalar multiples of each other (this is where the constraint qualification comes from).

Since x^* is a local maximum of $f(x)$ subject to $g(x) = 0$, we know that $f(x^*) \geq f(y)$ for all y in a neighborhood of x^*. Now, we compute the directional derivative of $\mathcal{L}(x^*, \lambda^*)$, considering only changes in x^* (we are looking for some profitable local deviation from x^*). This yields

$$D_y \mathcal{L}(x^*, \lambda^*) = \left(\nabla_x f(x^*) - \lambda^{*\prime} \nabla g(x^*) \right) \cdot y$$

But the term in parentheses equals zero by the first-order necessary conditions, so that there is no direction y in which the value of $\mathcal{L}(x^*, \lambda^*)$ increases.

Therefore, if x^* is a local maximum of $f(x)$ subject to $g(x) = 0$ and $\nabla g(x^*)$ is a non-singular matrix, then there exists λ^* such that

$$\nabla_x f(x^*) = \lambda^{*\prime} \nabla g(x^*)$$

and (x^*, λ^*) is a critical point of $\mathcal{L}(x, \lambda)$. ∎

Basically, the FONCs say that any local maximum of $f(x)$ subject to $g(x) = 0$ is a critical point of the Lagrangian when the Lagrangian is viewed as an unconstrained maximization problem in (x, λ). As usual, these are *necessary* conditions and help us identify a *candidate list*:

- Points where $f(x)$ and $g(x)$ are non-differentiable
- Points where the constraint qualification fails
- The critical points of the Lagrangian

All we know from the FONCs is that the global maximizer of $f(x)$ subject to $g(x) = 0$ must be on the list, not whether any particular entry on the list is a maximum, minimum, or saddle.

Example Consider the problem

$$\max_{x,y} xy$$

subject to $a = bx + cy$. The Lagrangian is

$$\mathcal{L}(x, y, \lambda) = xy - \lambda(bx + cy - a)$$

with FONCs

$$\begin{aligned} \mathcal{L}_x(x, y, \lambda) = y - \lambda b &= 0 \\ \mathcal{L}_y(x, y, \lambda) = x - \lambda c &= 0 \\ \mathcal{L}_\lambda(x, y, \lambda) = -(bx + cy - a) &= 0 \end{aligned}$$

To solve the system, notice that the first two equations can be rewritten as

$$\begin{aligned} y &= \lambda b \\ x &= \lambda c \end{aligned}$$

so that

$$\frac{y}{x} = \frac{b}{c}$$

Solving in terms of x yields

$$y = \frac{b}{c}x$$

We have one equation left, the constraint. Substituting the above equation into the constraint yields

$$a = bx + c\frac{b}{c}x$$

or

$$x^* = \frac{a}{2b}$$

$$y^* = \frac{a}{2c}$$

and

$$\lambda^* = \frac{a}{2bc}$$

Since $f(x)$ and $g(x)$ are continuously differentiable and the constraint qualification is everywhere satisfied, the candidate list consists of this single critical point (As of yet, we cannot determine whether it is a maximum or a minimum, but it's a maximum.). ∎

Example Consider the firm cost minimization problem: A firm hires capital K and labor L at prices r and w to product output $F(K, L) = q$. The firm minimizes cost subject to achieving output $\bar{q}$. Then the firm is trying to solve

$$\min_{K,L} rK + wL$$

subject to $F(K, L) = \bar{q}$. First, we convert to a maximization problem,

$$\max_{K,L} -rK - wL$$

subject to $g(K, L) = -F(K, L) + \bar{q} = 0$, and then form the Lagrangean,

$$\mathcal{L}(z, \lambda) = -rK - wL - \lambda(\bar{q} - F(K, L))$$

Then the FONC's are

$$-r + \lambda F_K(K^*, L^*) = 0$$

$$-w + \lambda F_L(K^*, L^*) = 0$$

Which characterizes the cost-minimizing plan in terms of w, r, and $\bar{q}$. ∎

Example Consider maximizing $f(x, y) = xy$ subject to the constraint $g(x, y) = x^2 + y^2 - 1$, so we are trying to make xy as large as possible on the unit circle. Then the Lagragnean is

$$\mathcal{L}(x, y, \lambda) = xy - \lambda(x^2 + y^2 - 1)$$

and the FONCs are

$$y^* - \lambda^* 2x^* = 0$$

$$x^* - \lambda^* 2y^* = 0$$

$$-(x^{*2} + y^{*2} - 1) = 0$$

Then we can use the same approach, re-arranging and dividing the first two equations. This yields

$$x^{*2} = y^{*2}$$

So the set of critical points are all the points on the circle such that $\sqrt{x^{*2}} = \sqrt{y^{*2}}$. There are four such points,

$$\left(\pm\sqrt{\frac{1}{2}}, \pm\sqrt{\frac{1}{2}}\right)$$

There are two global maxima, and two global minima. One maximum is

$$\left(\sqrt{\frac{1}{2}}, \sqrt{\frac{1}{2}}\right)$$

and the other is

$$\left(-\sqrt{\frac{1}{2}}, -\sqrt{\frac{1}{2}}\right)$$

The two global minima are

$$\left(+\sqrt{\frac{1}{2}}, -\sqrt{\frac{1}{2}}\right)$$

and

$$\left(-\sqrt{\frac{1}{2}}, +\sqrt{\frac{1}{2}}\right)$$

Right? ∎

10.2.1 The Geometry of Constrained Maximization

Lagrange maximization is actually very intuitive geometrically.

Let's start by recalling what *indifference curves* are. Suppose we fix a value, C, and set

$$f(x, y) = C$$

Then by the implicit function theorem, there is a function $x(y)$ that solves

$$f(x(y), y) = C$$

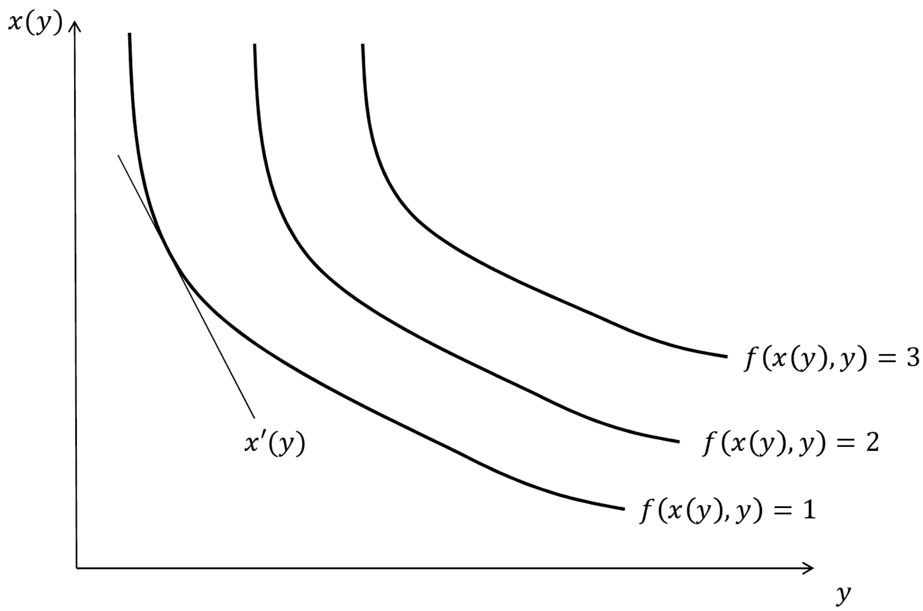
at least locally near y for some C. Now if we totally differentiate with respect to y, we get

$$f_x x'(y) + f_y = 0$$

and

$$\frac{\partial x(y)}{\partial y} = -\frac{f_y(x(y), y)}{f_x(x(y), y)}$$

This derivative is called the *marginal rate of substitution between x and y*: It expresses how much x must be given to or taken from the agent to compensate him for a small increase in the amount of y. So if we give the agent one more apple, he is presumably better off, so we have to take away a half a banana, and so on. If we graph $x(y)$ in the plane, we see the set of bundles (x, y) which all achieve an $f(x, y) = C$ level of satisfaction. The agent is *indifferent* among all these bundles, and the locus of points $(y, x(y))$ is an *indifference curve*.



Indifference Curves

Since $\nabla f(x, y) \geq 0$, the set of points above an indifference curve are all better than anything on the indifference curve, and we call these the *upper contour sets* of $f(x)$:

$$UC(a) = \{x : f(x) \geq a\}$$

while the set of points below an indifference curve are all worse than anything on the indifference curve, and we call these the *lower contour sets* of $f(x)$:

$$LC(a) = \{x : g(x) \leq a\}$$

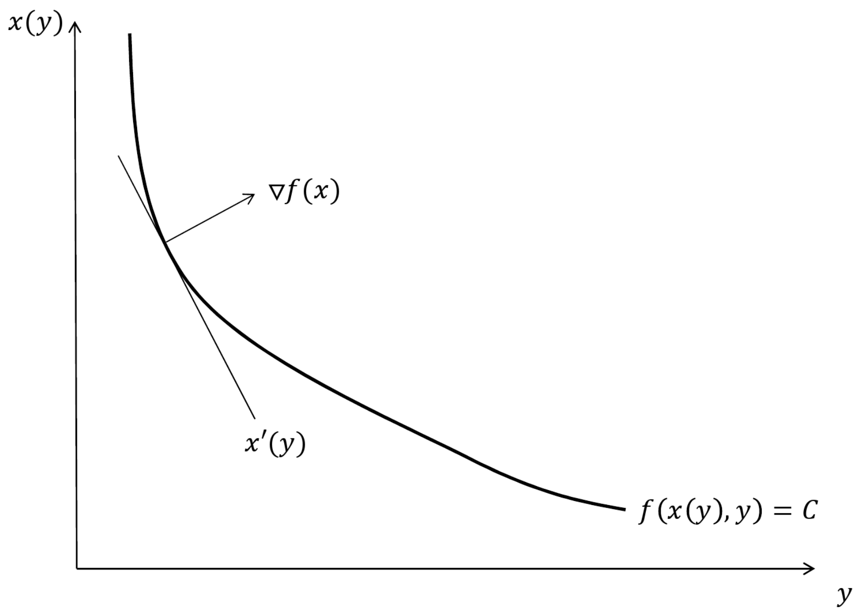
Now, along any indifference curve $(y, x(y))$, let's compute

$$\nabla f(x) \cdot (x'(y), 1)$$

This gives

$$-f_x \frac{f_y}{f_x} + f_y = 0$$

Recall that if $v_1 \cdot v_2 = 0$, then v_1 and v_2 are at a right-angle to each other (they are orthogonal). This implies that the gradient and the indifference curve are at right angles to one another.



Tangency of the Gradient to Indifference Curves

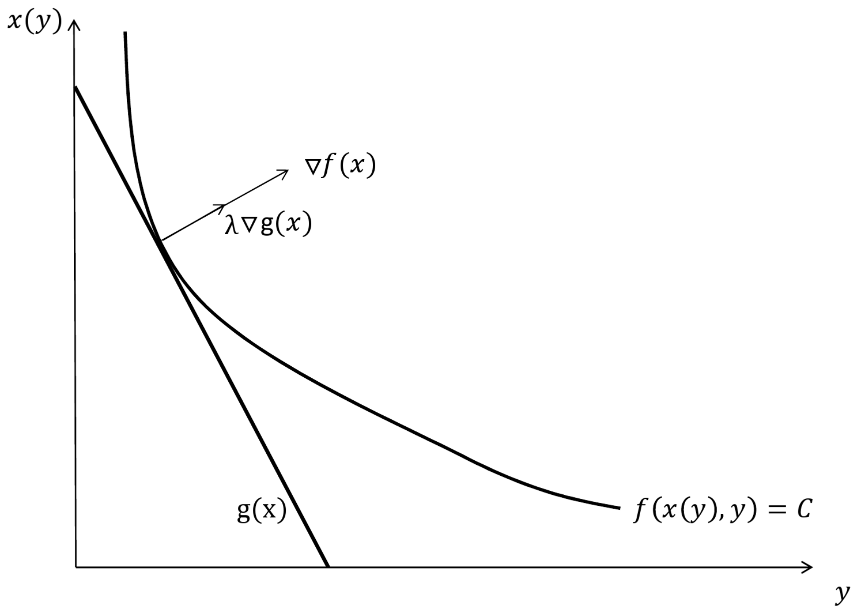
If $\nabla f(x) \geq 0$, the gradient gives the direction in which the function is increasing. Now, if we started on the interior of the constraint set and consulted the gradient, taking a step in the x direction if $f_x(x,y) \geq f_y(x,y)$ and taking a step in the y direction if $f_y(x,y) > f_x(x,y)$, we would drift up to the constraint. At this point, we would be on an indifference curve that is orthogonal to the gradient and tangent to the constraint, $g(x)$, and any more steps would violate the constraint.

Now, the FONCs are

$$\nabla f(x^*) = \lambda \nabla g(x^*)$$

$$g(x^*) = 0$$

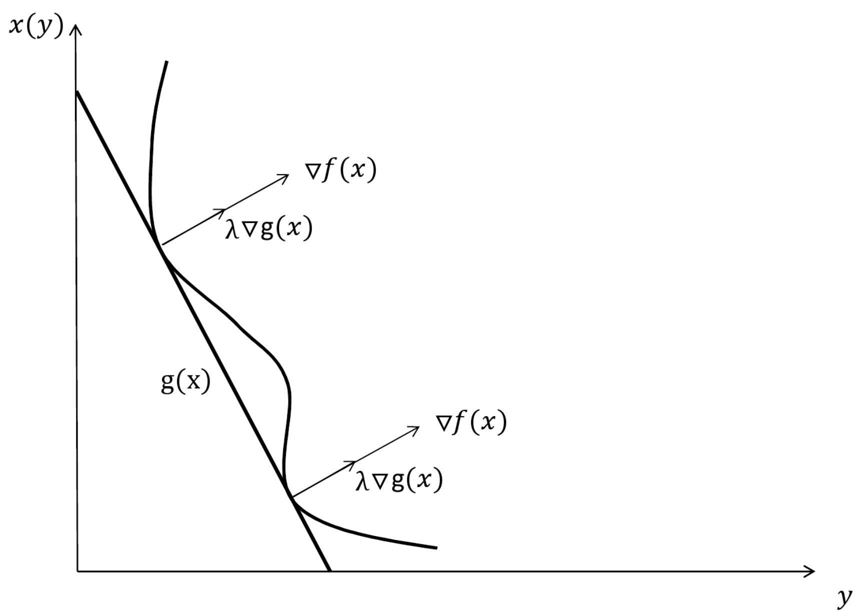
Recall that the gradient, $\nabla f(x)$, is a vector from the point $f(x)$ in the direction in which the function has the greatest rate of change. So the equation $\nabla f(x^*) = \lambda \nabla g(x^*)$ implies that the constraint gradient is a scalar multiple of the gradient of the objective function:



Tangency of the Gradient to the Constraint

So at a local maximum, it must be the case that the indifference curve is tangent to the constraint, since the constraint gradient points the same direction as the gradient of the objective, and the gradient of the objective is tangent to the indifference curve. This is all of the intuition: We're looking for a spot where the agent's indifference curve is tangent to the constraint, so that the agent can't find any local changes in which his payoff improves.

But if the function or constraint set have a lot of curvature, there can be multiple local solutions:



Multiple Solutions

This is the basic idea of Lagrange maximization, expressed graphically.

10.2.2 The Lagrange Multiplier

What is the interpretation of the Lagrange multiplier? Consider the problem of maximizing $f(x)$ subject to a linear constraint $w = p'x$, $p \gg 0$. You can think of w as an agent's wealth, p as the prices, and $f(x)$ as their payoff from consuming a bundle x. The Lagrangean evaluated at the optimum is:

$$\mathcal{L}(x^*, \lambda^*) = f(x^*) - \lambda(p'x^* - w)$$

But if we consider this instead as a function of w and differentiate, we get

$$V'(w) = \underbrace{(\nabla f(x^*(w)) - \lambda p)}_{\text{First-order condition}} \nabla_w x^*(w) - \frac{\partial \lambda^*}{\partial w} \underbrace{(p \cdot x^* - w)}_{=0} + \lambda^*(w)$$

So $V'(w) = \lambda^*(w)$, and the Lagrange multiplier gives the change in the optimized value of the objective for a small relaxation of the constraint set. Economists often call it the *shadow price of* w — how much the decision-maker would be willing to pay to relax the constraint slightly.

10.2.3 The Constraint Qualification

The constraint qualification can be confusing because it is a technical condition that has nothing to do with maximization. To make some sense of it (and generalize things a bit), consider the following problem:

$$\max_x f(x)$$

subject to $m = 1, 2, ..., M$ equality constraints, $g_1(x) = 0$, $g_2(x) = 0$, ..., $g_M(x) = 0$. Let

$$g(x) = \begin{bmatrix} g_1(x) \\ g_2(x) \\ \vdots \\ g_M(x) \end{bmatrix}$$

The new Langrangean is given by

$$\mathcal{L}(x, \lambda) = f(x) - \sum_{m=1}^{M} \lambda_m g_m(x) = f(x) - \lambda' g(x)$$

Theorem 10.2.3 (First-Order Necessary Conditions) *If*

1. *x^* is a local maximum of $f(x)$ that satisfies the constraints $g(x) = 0$*
2. *the constraint gradients $\nabla g(x^*) \neq 0$ (this is called the* constraint qualification*)*
3. *$f(x)$ and $g(x)$ are both differentiable at x^**

Then there exists a unique vector λ^* such that the FONCs

$$\nabla_x \mathcal{L}(x^*, \lambda^*) = \nabla f(x^*) - \lambda^{*\prime} \nabla g(x^*)$$

$$\nabla_\lambda \mathcal{L}(x^*, \lambda^*) = -g(x^*) = 0$$

hold.

Now, the new FONCs are

$$\nabla f(x) - \lambda' \nabla g(x) = \nabla f(x) - \sum_{m=1}^{M} \lambda_m \nabla g_m(x)$$

$$-g(x) = 0$$

The term

$$\nabla g(x) = \begin{bmatrix} \frac{\partial g_1(x)}{\partial x_1} & \frac{\partial g_1(x)}{\partial x_2} & \cdots & \frac{\partial g_1(x)}{\partial x_N} \\ \frac{\partial g_2(x)}{\partial x_1} & \frac{\partial g_2(x)}{\partial x_2} & & \\ \vdots & & \ddots & \\ \frac{\partial g_M(x)}{\partial x_1} & & & \frac{\partial g_M(x)}{\partial x_N} \end{bmatrix}$$

is actually an $M \times N$ matrix. Call this matrix G, just to clean things up a bit. Now our FONCs are

$$\nabla f(x) = \lambda' G$$

$$g(x) = 0$$

Remember that we are trying to solve for λ — that's what Lagrange's theorem guarantees — so we need G to be invertible, so that

$$\lambda^* = G^{-1} \nabla f(x)$$

But if G fails to be invertible, we can't solve for λ^*, and we can't "finish the proof".

So the constraint qualification is just saying, "The matrix G is invertible," nothing more.

Example Here's an example where the constraint qualification fails with only one constraint:

$$\max_{\{x,y: y^3 - x^2 = 0\}} -y$$

The Lagrangean is

$$\mathcal{L} = -y - \lambda(y^3 - x^2)$$

with first-order necessary conditions

$$-(y^3 - x^2) = 0$$

$$-1 - \lambda 3y^2 = 0$$

$$\lambda 2x = 0$$

The constraint $y^3 - x^2 = 0$ requires that both x and y be (weakly) positive. Since the objective is decreasing in y and non-responsive in x, the global maximum is at $y^* = x^* = 0$. But then the first-order necessary conditions are

$$\begin{bmatrix} 0 \\ -1 \\ 0 \end{bmatrix} \neq \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

and cannot be satisfied. ∎

So we have a simple situation where the Lagrange approach fails. The reason is that the constraints have gradient

$$\begin{bmatrix} \partial g(0,0)/\partial x \\ \partial g(0,0)/\partial y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

so the constraint vanishes at the optimum. Then $\nabla \mathcal{L} = \nabla f + \lambda \nabla g$ cannot be solved if $\nabla f \neq 0$ but ∇g is. This is the basic idea of the constraint qualification.

10.3 Second-Order Sufficient Conditions

Just like unconstrained problems, there are second-order conditions, but they more complicated and require patience and practice to understand. There are two main differences from the unconstrained problem.

First, we need not consider *all* possible local changes in the choice variables x^* to verify whether a particular critical point is a local maximum, but *only those that do not violate the constraints*. For example, if the function if $f(x, y) = xy$ and the constraint is $w = px + y$, there are always better allocations nearby any critical point, but most of them violate the constraint $w = px + y$. The set of changes (dx, dy) that leave the total expenditure unchanged satisfy $pdx + dy = 0$, and these are the only changes which we are really interested in.

Definition 10.3.1 *Let (x^*, λ^*) be a critical point of the Lagrangian. The set of* feasible local variations *is*

$$\mathcal{Y} = \{y \in \mathbb{R}^n : \nabla g(x^*) \cdot y = 0\}$$

This is the set of vectors y local to x^* for which all of the constraints are still satisfied, in the sense of the directional derivative, $\nabla g(x^*) \cdot y = 0$.

Second, the Hessian of $\mathcal{L}(x, \lambda)$ is not just the Hessian of $f(x)$. Intuitively, I have suggested that we should think of equality constrained maximization of x subject to $g(x) = 0$ as unconstrained maximization of $\mathcal{L}(x, \lambda)$ in terms of (x, λ). This is reflected in the fact that the Hessian of $\mathcal{L}(x, \lambda)$ is the focus of the SOSCs, not the Hessian of $f(x)$.

Theorem 10.3.2 (Second-Order Sufficient Conditions) *Suppose (x^*, λ^*) is a critical point of the Lagrangian (i.e., the constraint qualification and the FONCs $\nabla f(x^*) + \lambda^{*\prime} \nabla g(x^*) = 0$ and $g(x^*) = 0$ are satisfied). Let $\nabla_{xx}\mathcal{L}(x^*, \lambda^*)$ be the Hessian of the Lagrangean function with respect to the choice variables only. Then*

- *If $y' \nabla_{xx}\mathcal{L}(x^*, \lambda^*) y < 0$ for all feasible local changes y, then x^* is a local maximum of f*
- *If $y' \nabla_{xx}\mathcal{L}(x^*, \lambda^*) y > 0$ for all feasible local changes y, then x^* is a local minimum of f*

Since we are not solving an unconstrained maximization problem and the solution must lie on the locus of points defined by $g(x) = 0$, it follows that we can imagine checking the local, second-order sufficient conditions only on the locus. Since we are using calculus, we restrict attention to feasible local changes, and the consequence is that for all "small steps" y that do not violate the constraint, the Lagrangian must have a negative definite Hessian in terms of x alone. This does not mean that the Hessian of the Lagrangian is negative definite, because the Lagrangian depends on x and λ. However, we would like a test that avoids dealing with the set of feasible local variations directly, since that set appears difficult to solve for and manipulate.

Example If

$$\mathcal{L}(x, y, \lambda) = f(x, y) - \lambda(ax + by - c)$$

then the FONCS are

$$-(ax + by - c) = 0$$
$$f_x(x^*, y^*) - \lambda^* a = 0$$
$$f_y(x^*, y^*) - \lambda^* b = 0$$

Now, differentiating all the equations again in the order λ^*, x^*, and y^* gives the matrix

$$\nabla_{xx}\mathcal{L}(x^*, y^*, \lambda^*) = \begin{bmatrix} 0 & -a & -b \\ -a & f_{xx}(x^*, y^*) & f_{yx}(x^*, y^*) \\ -b & f_{xy}(x^*, y^*) & f_{yy}(x^*, y^*) \end{bmatrix}$$

The above Hessian is called *bordered*, since the first column and first row are transposes of each other, while the lower right-hand corner is the Hessian of the objective function.

For a single constraint $g(x)$, the Hessian of the Lagrangean looks like this:

$$\nabla_{xx}\mathcal{L}(x^*,\lambda^*) = \begin{bmatrix} 0 & \nabla_x g(x^*)' \\ \nabla_x g(x^*) & \nabla_{xx} f(x^*) + \lambda^* \nabla_{xx} g(x^*) \end{bmatrix} \tag{10.1}$$

This is called the *bordered Hessian.* ∎

Suppose you are maximizing $f(x)$ subject to a single constraint, $g(x) = 0$ (you can find the generalization to any number of constraints in any math econ textbook). Then the determinant of the upper-left hand corner is trivially zero. The *non-trivial leading principal minors of the bordered Hessian* are — for the case with a single constraint —

$$H_3 = \begin{bmatrix} 0 & -g_{x_1}(x^*) & -g_{x_2}(x^*) \\ -g_{x_1}(x^*) & f_{x_1x_1}(x^*) - \lambda^* g_{x_1x_1}(x^*) & f_{x_2x_1}(x^*) - \lambda^* g_{x_2x_1}(x^*) \\ -g_{x_2}(x^*) & f_{x_1x_2}(x^*) - \lambda^* g_{x_1x_2}(x^*) & f_{x_2x_2}(x^*) - \lambda^* g_{x_2x_2}(x^*) \end{bmatrix}$$

$$H_4 = \begin{bmatrix} 0 & -g_{x_1}(x^*) & -g_{x_2}(x^*) & -g_{x_3}(x^*) \\ -g_{x_1}(x^*) & f_{x_1x_1}(x^*) - \lambda^* - g_{x_1x_1}(x^*) & f_{x_2x_1}(x^*) - \lambda^* - g_{x_2x_1}(x^*) & f_{x_3x_1}(x^*) - \lambda^* - g_{x_3x_1}(x^*) \\ -g_{x_2}(x^*) & f_{x_1x_2}(x^*) - \lambda^* - g_{x_1x_2}(x^*) & f_{x_2x_2}(x^*) - \lambda^* - g_{x_2x_2}(x^*) & f_{x_3x_2}(x^*) - \lambda^* - g_{x_3x_2}(x^*) \\ -g_{x_3}(x^*) & f_{x_1x_3}(x^*) - \lambda^* - g_{x_1x_3}(x^*) & f_{x_2x_3}(x^*) - \lambda^* - g_{x_2x_3}(x^*) & f_{x_3x_3}(x^*) - \lambda^* - g_{x_3x_3}(x^*) \end{bmatrix}$$

$$\vdots$$

$$H_{k+1} = \begin{bmatrix} 0 & -g_{x_1}(x^*) & \dots & -g_{x_k}(x^*) \\ -g_{x_1}(x^*) & f_{x_1x_1}(x^*) - \lambda^* g_{x_1x_1}(x^*) & & \\ \vdots & & \ddots & \\ -g_{x_k}(x^*) & & & f_{x_kx_k}(x^*) - \lambda^* g_{x_kx_k}(x^*) \end{bmatrix}$$

So H_{k+1} is the upper left-hand $k+1 \times k+1$ principal minor of the full bordered Hessian. The "+1" term comes from the fact that we have an extra leading row and column that correspond to the constraint.

An example makes this a bit clearer:

Example Let

$$f(x_1, x_2, x_3) = x_1x_2 + x_2x_3 + x_1x_3$$

subject to

$$x_1 + x_2 + x_3 = 3$$

Then the Lagrangean is

$$\mathcal{L} = x_1x_2 + x_2x_3 + x_1x_3 - \lambda(x_1 + x_2 + x_3 - 3)$$

This generates a system of first-order conditions:

$$\begin{bmatrix} \partial\mathcal{L}/\partial\lambda \\ \partial\mathcal{L}/\partial x_1 \\ \partial\mathcal{L}/\partial x_2 \\ \partial\mathcal{L}/\partial x_3 \end{bmatrix} = \begin{bmatrix} (3 - x_1 - x_2 - x_3) \\ x_2 + x_3 - \lambda \\ x_1 + x_3 - \lambda \\ x_2 + x_1 - \lambda \end{bmatrix}$$

The only critical point is $x^* = (1/3, 1/3, 1/3)$. The bordered Hessian at the critical point is

$$\nabla_{xx}\mathcal{L}(x^*,\lambda^*) = \begin{bmatrix} 0 & -1 & -1 & -1 \\ -1 & 0 & 1 & 1 \\ -1 & 1 & 0 & 1 \\ -1 & 1 & 1 & 0 \end{bmatrix}$$

and the leading principal minors are

$$H_3 = \begin{bmatrix} 0 & -1 & -1 \\ -1 & 0 & -1 \\ -1 & -1 & 0 \end{bmatrix}$$

and

$$H_4 = \begin{bmatrix} 0 & -1 & -1 & -1 \\ -1 & 0 & 1 & 1 \\ -1 & 1 & 0 & 1 \\ -1 & 1 & 1 & 0 \end{bmatrix}$$

$\blacksquare$

It turns out that we can develop a test based on the bordered Hessian of the whole Lagrangian, including derivatives with respect to the multiplier, rather than focusing on the Hessian of the Lagrangian restricted only to the choice variables.

Theorem 10.3.3 (Alternating Sign Test)

- *A critical point (x^*, λ^*) is a local maximum of $f(x)$ subject to the constraint $g(x) = 0$ if the determinants of all the principal minors of the bordered Hessian alternate in sign, starting with*

$$\det H_3 > 0$$

- *A critical point (x^*, λ^*) is a local minimum of $f(x)$ subject to the constraint $g(x) = 0$ if the determinants of all the principal minors of the bordered Hessian are all negative, starting with*

$$\det H_3 < 0$$

Returning to the example,

Example We need to decide whether the bordered Hessian is negative definite or not. The non-trivial leading principal minors of the bordered Hessian are

$$det(H_3) = 2 > 0, det(H_4) = -3 < 0$$

And the second-order condition is satisfied for any vector $y'Hy$. $\blacksquare$

Example Let

$$f(x_1, x_2) = f_1(x_1) + f_2(x_2)$$

where $f_1(x_1)$ and $f_2(x_2)$ are increasing and $f_1''(x), f_2''(x) < 0$, and $f_i(0) = 0$. There is a constraint that $C = x_1 + x_2$. Then the Lagrangean is

$$\mathcal{L} = f_1(x_1) + f_2(x_2) - \lambda(x_1 + x_2 - C)$$

This generates a system of first-order conditions

$$\begin{bmatrix} \partial\mathcal{L}/\partial\lambda \\ \partial\mathcal{L}/\partial x_1 \\ \partial\mathcal{L}/\partial x_2 \end{bmatrix} = \begin{bmatrix} -(x_1 + x_2 - C) \\ f_1'(x_1) - \lambda \\ f_2'(x_2) - \lambda \end{bmatrix}$$

with bordered Hessian

$$H(x, \lambda) = \begin{bmatrix} 0 & -1 & -1 \\ -1 & f_1''(x_1) & 0 \\ -1 & 0 & f_2''(x_2) \end{bmatrix}$$

Then the bordered Hessian's non-trivial leading principal minors starting from H_3 have determinants

$$det(H_3) = -f_1''(x_1) - f_2''(x_2) > 0$$

So the Lagrangian is negative definite, and an interior solution to the first-order conditions is a solution to the maximization problem. ∎

Example Suppose a consumer is trying to maximize utility, with utility function $u(x, y) = x^\alpha y^\beta$ and budget constraint $w = px + y$. Then we can write the Lagrangian as (why?)

$$\mathcal{L} = \alpha \log(x) + \beta \log(y) - \lambda(px + y - w)$$

we get a system of first-order conditions

$$\begin{bmatrix} \partial\mathcal{L}/\partial\lambda \\ \partial\mathcal{L}/\partial x \\ \partial\mathcal{L}/\partial y \end{bmatrix} = \begin{bmatrix} w - px - y \\ \alpha/x - \lambda p \\ \beta/y - \lambda \end{bmatrix}$$

The Hessian of the Lagrangean is

$$\begin{bmatrix} 0 & -p & -1 \\ -p & \dfrac{-\alpha}{(x^*)^2} & 0 \\ -1 & 0 & \dfrac{-\beta}{(y^*)^2} \end{bmatrix}$$

The determinants of the non-trivial leading principal minors of the bordered Hessian are

$$\det \begin{bmatrix} 0 & -p & -1 \\ -p & \dfrac{-\alpha}{(x^*)^2} & 0 \\ -1 & 0 & \dfrac{-\beta}{(y^*)^2} \end{bmatrix} = p(-p)\frac{-\beta}{(x^*)^2} - 1(-1)(-1)\frac{-\alpha}{(x^*)^2} > 0$$

So the alternating sign test is satisfied. ∎

This works because adding the border imposes the restriction that $\nabla g(x) = 0$ on the test of whether the objective is negative definite or not. Consequently, when we use the alternating sign test, we are really asking, "Is the Hessian of the objective negative definite on Y?" This requires more matrix algebra to show, but you can find the details in MWG or Debreu's papers, for example.

10.4 Comparative Statics

Our system of first-order necessary conditions

$$\nabla_x f(x^*) - \lambda^* \nabla_x g(x^*) = 0$$

$$-g(x^*) = 0$$

is a non-linear system of equations with endogenous variables (x^*, λ^*), just like any other we have applied the IFT to. The only "new" part is that you have to keep in mind that λ^* is an endogenous variable, so that it changes when we vary any exogenous variables. Second, the sign of the determinant of the Hessian is determined by whether we are looking at a maximum or minimum, and the number of equations.

For simplicity, consider a constrained maximization problem

$$\max_{x,y} f(x,y,t)$$

subject to a single equality constraint $ax + by - s = 0$, where t and s are exogenous variables. This is simpler than assuming that $g(x,s) = 0$, since the second-order derivatives of the constraint are all zero. However, there are many economic problems of interest where the constraints are linear, and a general $g(x,s)$ can be very complicated to work with. Then the Lagrangean is

$$\mathcal{L}(x,y,\lambda) = f(x,y,t) - \lambda(ax + by - s)$$

and the FONCs are

$$-(ax^* + by^* - s) = 0$$

$$\nabla f(x^*, y^*, t) - \lambda^* \nabla g(x^*, y^*, s) = 0$$

Since we have two parameters — t shifts the objective function, s shifts the constraint — we can look at two different comparative statics.

Example Let's start with t: Totally differentiate the FONCs to get three equations

$$-\left(a\frac{\partial x^*}{\partial t} + b\frac{\partial y^*}{\partial t}\right) = 0$$

$$f_{tx} + f_{xx}\frac{\partial x^*}{\partial t} + f_{yx}\frac{\partial y^*}{\partial t} - \frac{\partial \lambda^*}{\partial t}a = 0$$

$$f_{ty} + f_{xy}\frac{\partial x^*}{\partial t} + f_{yy}\frac{\partial y^*}{\partial t} - \frac{\partial \lambda^*}{\partial t}b = 0$$

If we write this as a matrix equation,

$$\begin{bmatrix} 0 & -a & -b \\ -a & f_{xx} & f_{xy} \\ -b & f_{yx} & f_{yy} \end{bmatrix} \begin{bmatrix} \partial \lambda^*/\partial t \\ \partial x^*/\partial t \\ \partial y^*/\partial t \end{bmatrix} = \begin{bmatrix} 0 \\ -f_{tx} \\ -f_{ty} \end{bmatrix}$$

Which is just a regular "$Ax = b$"-type matrix equation. On the right-hand side, the bordered Hessian appears; from the alternating sign test and the fact that x^* is a local maximum, we can determine the sign of its determinant. To some extent, we are finished, since all that remains is to solve the system. You can solve the system by hand by solving for each comparative static and reducing the number of equations, but that is a lot of work. It is simpler to use Cramer's rule. To see how x^* varies with t,

$$\frac{\partial x^*}{\partial t} = \frac{\det \begin{bmatrix} 0 & 0 & -b \\ -a & -f_{tx} & f_{xy} \\ -b & -f_{ty} & f_{yy} \end{bmatrix}}{\det H_3}$$

which is

$$\frac{\partial x^*}{\partial t} = \frac{-b(af_{ty}(x^*, y^*, t) - bf_{tx}(x^*, y^*, t)}{\det H_3}$$

Since (x^*, y^*) is a local maximum, $\det H_3 > 0$, and the sign of the comparative static is just the sign of the numerator. $\blacksquare$

Having done t, let's do s:

Example Totally differentiate the system of FONCs to get

$$-\left(a\frac{\partial x^*}{\partial s}+b\frac{\partial y^*}{\partial s}-1\right)=0$$

$$f_{xx}\frac{\partial x^*}{\partial s}+f_{yx}\frac{\partial y^*}{\partial s}-\frac{\partial \lambda^*}{\partial s}a=0$$

$$f_{xy}\frac{\partial x^*}{\partial s}+f_{yy}\frac{\partial y^*}{\partial s}-\frac{\partial \lambda^*}{\partial s}b=0$$

Writing this as a matrix equation yields

$$\begin{bmatrix} 0 & -a & -b \\ -a & f_{xx} & f_{xy} \\ -b & f_{yx} & f_{yy} \end{bmatrix}\begin{bmatrix} \partial\lambda^*/\partial s \\ \partial x^*/\partial s \\ \partial y^*/\partial s \end{bmatrix}=\begin{bmatrix} -1 \\ 0 \\ 0 \end{bmatrix}$$

Let's focus on $\partial y^*/\partial s$. Using Cramer's rule,

$$\frac{\partial y^*}{\partial s}=\frac{\det\begin{bmatrix} 0 & -a & -1 \\ -a & f_{xx} & 0 \\ -b & f_{yx} & 0 \end{bmatrix}}{\det H_3}$$

$$\frac{\partial y^*}{\partial s}=\frac{af_{yx}+bf_{xx}}{\det H_3}$$

■

An example with more economic content is:

Example Consider the consumer problem

$$\mathcal{L}=\alpha\log(x)+\beta\log(y)-\lambda(px+y-w)$$

we get a system of first-order conditions

$$\begin{bmatrix} \partial\mathcal{L}/\partial\lambda \\ \partial\mathcal{L}/\partial x \\ \partial\mathcal{L}/\partial y \end{bmatrix}=\begin{bmatrix} w-px-y \\ \alpha/x-\lambda p \\ \beta/y-\lambda \end{bmatrix}$$

This is a system of non-linear equations, so there's no problem differentiating it with respect to p, for instance, and using all the same steps as for the unconstrained firm. You need to remember, however, that $\lambda^*(w,p,\alpha,\beta)$ is now a function of parameters as well, so it is not a constant that drops out. Totally differentiating with respect to p gives

$$\begin{aligned} 0 &= -x-p\frac{\partial x}{\partial p}-\frac{\partial y}{\partial p} \\ 0 &= \frac{-\alpha}{x^2}\frac{\partial x}{\partial p}-\frac{\partial\lambda}{\partial p}p-\lambda \\ 0 &= \frac{-\beta}{y^2}\frac{\partial y}{\partial p}-\frac{\partial\lambda}{\partial p} \end{aligned}$$

Solving the system the "long way" for $\partial x/\partial p$ gives

$$\frac{\partial x}{\partial p}=-\frac{\frac{\beta}{y^{*2}}px^*+\lambda^*}{\frac{\alpha}{x^{*2}}+\frac{\beta}{y^{*2}}p^2}<0$$

Since the Lagrange multiplier λ^* is positive, if p goes up, the consumer reduces his consumption of x^*. ■

10.5 The Envelope Theorem

The last tool we need to generalize is the envelope theorem. Since a local maximum of the Lagrangian satisfies

$$\mathcal{L}(x^*, \lambda^*) = f(x^*) - \lambda^{*\prime} g(x^*) = f(x^*)$$

we can use it as our value function.

As a start, suppose we are studying the consumer's problem with utility function $u(x, y)$ and budget constraint $w = px + y$. Then the Lagrangian at any critical point is

$$\mathcal{L}(x^*, y^*, \lambda^*) = u(x^*, y^*) - \lambda^*(px^* + y^* - w)$$

Let's consider it as a value function in terms of w,

$$V(w) = u(x^*(w), y^*(w)) - \lambda^*(w)(px^*(w) + y^*(w) - w)$$

Then

$$V'(w) = \underbrace{u_x x^{*\prime}(w) + u_y y^{*\prime}(w)}_{\text{I}} \underbrace{-\lambda^*(w)px^{*\prime}(w) - \lambda^*(w)y^{*\prime}(w)}_{\text{II}} \underbrace{-\lambda^{*\prime}(w)(px^*(w) + y^*(w) - w)}_{\text{III}} \underbrace{+\lambda^*(w)}_{\text{IV}}$$

There are a bunch of terms that each represent a different consequence of giving the agent more wealth. First, the agent re-optimizes and the value of the objective function changes (I). Second, this has an impact on the constraint, and its implicit cost changes (II). Third, since the constraint has been relaxed, the Lagrange multiplier changes, again changing the implicit cost (III). Fourth, there is the direct effect on the objective of increasing w (IV).

Re-arranging, however, gives

$$V'(w) = \underbrace{(u_x - \lambda^*(w)p)x^{*\prime}(w) + (u_y - \lambda^*(w))y^{*\prime}(w)}_{\text{FONCs}} - \lambda^{*\prime}(w) \underbrace{(px^*(w) + y^*(w) - w)}_{\text{Constraint}} + \lambda^*(w)$$

so that the FONCs are zero, and since the constraint is satisfied, that term also equals zero, leaving just the direct effect,

$$V'(w) = \lambda^*(w)$$

or

$$V'(w) = \frac{\partial \mathcal{L}(x^*(w), y^*(w), \lambda^*(w), w)}{\partial w} = \frac{d\mathcal{L}(x^*(w), y^*(w), \lambda^*(w), w)}{dw}$$

So as before, we can simply take the partial derivative of the Lagrangean to see how an agent's payoff changes with respect to a parameter, rather than working out all the comparative statics through the implicit function theorem.

Theorem 10.5.1 (Envelope Theorem) *Consider the constrained maximization problem*

$$\max_x f(x, t)$$

subject to $g(x, s) = 0$. *Let* $x^*(t, s), \lambda^*(t, s)$ *be a critical point of the Lagrangean. Define*

$$V(t, s) = \mathcal{L}(x^*(t, s), \lambda^*(t, s), t, s) = f(x^*(t, s), t) + \lambda^*(t, s) g(x^*(t, s), s)$$

Then

$$\frac{\partial V(t, s)}{\partial t} = \frac{\partial f(x^*(t, s), t)}{\partial t}$$

and

$$\frac{\partial V(t, s)}{\partial s} = \frac{\partial g(x^*(t, s), s)}{\partial s}$$

In words, the envelope theorem implies that a change in the agent's payoff with respect to an exogenous variable is the partial derivative of the *non-optimized* Lagranian evaluated at the optimal decision $x^*(t, s)$.

Example Suppose a firm has production function $F(K, L, t)$ where K is capital, L is labor, and t is technology. The price of the firm's good is p. Then the we can write the profit maximization problem as a constrained maximization problem as

$$\pi(t) = \max_{q,K,L} pq - rK - wL$$

subject to $q = F(K, L, t)$. The Lagrangean then is

$$\mathcal{L}(q, K, L, \lambda_p) = pq - rK - wL - \lambda_p(q - F(K, L, t))$$

And — without any actual work —

$$\pi'(t) = \lambda_p^* F_t(K^*, L^*, t)$$

Similarly, the cost minimization problem is

$$C(q, t) = \min_{K,L} rK + wL$$

subject to $F(K, L, t) \geq q$. The Lagrangean then is

$$\mathcal{L}(K, L, \lambda) = -rK - wL + \lambda_c(F(K, L, t) - q)$$

And

$$C_t(q, t) = \lambda_c^* F_t(K^*, L^*, t)$$

$\blacksquare$

So be careful: The envelope theorem takes the derivative of the Lagrangean, then evaluates it at the optimal choice to obtain the derivative of the value function.

Exercises

1. Suppose we take a strictly increasing transformation of the objective function and leave the constraints unchanged. Is a solution of the transformed problem a solution of the original problem? Suppose we have constraints $g(x) = c$ and take a strictly increasing transformation of both sides. Is a solution of the transformed problem a solution of the original problem?

2. Consider the maximization problem

$$\max_{x_1,x_2} x_1 + x_2$$

subject to

$$x_1^2 + x_2^2 = 2$$

Sketch the constraint set and contour lines of the objective function. Find all critical points of the Lagrangian. Verify whether each critical point is a local maximum or a local minimum. Find the global maximum.

3. Consider the maximization problem

$$\max_{x,y} xy$$

subject to

$$\frac{a}{2}x^2 + \frac{b}{2}y^2 = r$$

Sketch the constraint set and contour lines of the objective function. Find all critical points of the Lagrangian. Verify whether each critical point is a local maximum or a local minimum. Find the global maximum. How does the value of the objective vary with r? a? How does x^* respond to a change in r; show this using the closed-form solutions and the IFT.

4. Solve the problem

$$\max_{x,y,z} 2x - 2y + z$$

subject to $x^2 + y^2 + z^2 = 9$.

5. Solve the following two dimensional maximization problems subject to the linear constraint $w = p_1x_1 + p_2x_2$, $p_1, p_2 > 0$. Then compute the change in x_1 with respect to p_2, and a change in x_1 with respect to w. Assume $\alpha_1 > \alpha_2 > 0$. For i and ii, when do the SOSCs hold?

i. Cobb-Douglas

$$f(x) = x_1^{\alpha_1} x_2^{\alpha_2}$$

ii. Stone-Geary

$$f(x) = (x_1 - \gamma_1)^{\alpha_1}(x_2 - \gamma_2)^{\alpha_2}$$

iii. Constant Elasticity of Substitution

$$\left(\alpha_1 x_1^{1/\rho} + \alpha_2 x_2^{1\rho}\right)^{\rho}$$

iv. Leontief

$$\min\{\alpha_1 x_1, \alpha_2 x_2\}$$

6. An agent receives income w_1 in period one and no income in the second period. He consumes c_1 in the first period and c_2 in the second. He can save in the first period, s, and receives back Rs in the second period. His utility function is

$$u(c_1, c_2) = \log(c_1) + \delta \log(c_2)$$

where $R > 1$ and $0 < \delta < 1$. (i) Write this down as a constrained maximization problem. (ii) Find first-order necessary conditions for optimality and check the SOSCs. What does the Lagrange multiplier represent in this problem? (iii) Compute the change in c_2 with respect to R and δ. (iv) How does the agent's payoff change if R goes up? δ?

7. For utility function

$$u(x_1, x_2) = (x_1 - \gamma_1)x_2^{\alpha}$$

and budget constraint $w = p_1x_1 + p_2x_2$, compute the utility-maximizing bundle. You should get closed-form solutions. Now compute $\partial x_1^*/\partial p_2$ using the closed-form solution as well as the implicit function theorem from the system of FONCs. Repeat with $\partial x_2^*/\partial p_1$. How does the agent's payoff vary in γ_1? w? p_2?

8. Generalize and proof the envelope theorem in the case where $f(x,t)$ and $g(x,t)$ both depend on the same parameter, t. Construct an economically relevant problem in which this occurs, and use the implicit function theorem to show how $x^*(t)$ varies in t when both the constraint and objective are varying at the same time.

9. Suppose you are trying to maximize $f(x)$. Show that if the constraint is linear, so that $p \cdot x = w$, you can always rewrite the constraint as

$$x_N = \frac{\sum_{i=1}^{N-1} p_i x_i}{p_N}$$

and substitute this into the objective, turning the problem into an unconstrained maximization problem. Why do we need Lagrange maximization then?

10. Suppose you have a maximization problem

$$\max_x f(x)$$

subject to $g(x, t) = 0$. Show how to use the implicit function theorem to derive comparative statics with respect to t. Explain briefly what the bordered Hessian looks like, and how it differs from the examples in the text.

Chapter 11

Inequality-Constrained Maximization Problems

Many problems — particularly in macroeconomics — don't involve strict equality constraints, but involve inequality constraints. For example, a household might be making a savings/consumption decision, subject to the constraint that savings can never become negative — this is called a *borrowing constraint*. As a result, many households will not have to worry about this, as long as they maintain positive savings in all periods. A simple version is:

Example Consider a household with utility function

$$u(c_1) + \delta u(c_2)$$

The household gets income y_1 in period one and y_2 in period two, with $y_2 > y_1$. The household would potentially like to borrow money today to smooth consumption, but faces a financial constraint that it can borrow an amount B only up to $R < y_2$. Then we have the constraints

$$c_1 = y_1 + B$$

$$c_2 = y_2 - B$$

$$B \leq R$$

Substituting the first two constraints into the objective yields

$$\max_B u(y_1 + B) + \delta u(y_2 - B)$$

subject to $B \leq y_2$.

Now, sometimes the constraint will be irrelevant, or

$$u'(y_1 + B^*) = \delta u'(y_2 - B^*)$$

and $B^* < R$, and the constraint is "slack," or "inactive." Other times, however, $B^* > R$, and the constraint becomes "binding" or "active". Then $B' = R$, and we have two potential solutions, and the one selected depends on the parameters of the problem. ∎

The simple borrowing problem illustrates the general issues: Having inequality constraints creates the possibility that for different values of the parameters, different collections of the constraints are binding. The maximization theory developed here is designed to deal with these shifting sets of constraints.

Definition 11.0.2 *Let $f(x)$ be the objective function. Suppose that the choice of x is subject to $m = 1, 2, ..., M$* equality constraints*, $g_m(x) = 0$, and $k = 1, 2, ..., K$* inequality constraints*, $h_k(x) \leq 0$. Then the* inequality-constrained maximization problem *is*

$$\max_x f(x)$$

such that for $m = 1, 2, ..., M$ and $k = 1, 2, ..., K$,

$$g_m(x) = 0$$

$$h_k(x) \leq 0$$

Note that we can brute force solve this problem as follows: Since there are K inequality constraints, there are 2^K ways to pick a subset of the inequality constraints. For each subset, we can make the chosen inequality constraints into equality constraints, and solve the resulting equality constrained maximization problem (which we already know how to do). Each of these sub-problems might generate no candidates, or many, and some of the candidates may conflict with some of the constraints we are ignoring. Once we compile a list of all the solutions that are actually feasible, then the global maximum must be on the list somewhere. Then we can simply compare the payoffs from our 2^K sets of candidate solutions, and pick the best.

Example Suppose an agent is trying to maximize $f(x, y) = ax + by$, $a, b > 0$, subject to the constraints $x \geq 0$, $y \geq 0$, and $c \geq px + qy$.

Since $\nabla f = (a, b) \gg 0$, it is immediate that the constraint $c \geq px+qy$ will bind with an equality. For if $c > px + qy$, we can always increase x or y a little bit, thus improving the objective function's value as well as not violating the constraint. This is feasible and better than any proposed solution with $c > px + qy$, so we must have $c = px + qy$.

Then we have three cases remaining:

1. The constraint $x \geq 0$ binds and $x = 0$, but $y > 0$
2. The constraint $y \geq 0$ binds and $y = 0$, but $x > 0$
3. The constraints $x, y \geq 0$ are both slack, and $x, y > 0$

We then solve case-by-case:

1. If $y = 0$, then the constraint implies $x = c/p$, giving a value of ac/p
2. If $x = 0$, then the constraint implies $y = c/q$, giving a value of bc/q
3. If $x, y > 0$, then we solve the constrained problem using a Lagrangean,

   $$\mathcal{L}(x, y, \lambda) = ax + by - \lambda(px + qy - c)$$

   with FONCs

   $$a - \lambda p = 0$$

   $$b - \lambda q = 0$$

   and

   $$-(px + qy - c) = 0$$

   This has a solution only if $a/b = p/q$, in which case any x and y that satisfy the constraints are a solution, since the objective is equivalent to $(a/b)x + y$ and the constraint is equivalent to $(p/q)x + y$, from which you can easily see that the indifference curves and constraint set are exactly parallel, so that *any* $px^* + qy^* = c$ with $x^*, y^* > 0$ is a solution.

So we have two candidate solutions with either x or y equal to zero, and a continuum of solutions when the objective and constraint are parallel. ∎

As happened in the previous problem, we differentiate between two kinds of solutions:

Definition 11.0.3 *A point x^* that is a local maximizer of $f(x)$ subject to $g_m(x) = 0$ for all m and $h_k(x) \leq 0$ for all k is a* corner solution *if $h_k(x^*) = 0$ for some k, and an* interior solution *if $h_k(x^*) < 0$ for all k.*

A corner solution is one like $x^* = c/p$ and $y^* = 0$, while an interior solution corresponds to the third case from the previous example, where the constraint $c = px + qy$ binds.

This example suggests we don't really need a new maximization theory, but either a lot of paper or a powerful computer. However, just as Lagrange multipliers can be theoretically useful, the multipliers we'll attach to the inequality constraints can be theoretically useful. In fact, the framework we'll develop is mostly just a systematic way of doing the brute force approach described in the previous paragraph.

11.1 First-Order Necessary Conditions

For simplicity, we'll work with one equality constraint $g(x) = 0$ and $k = 1, ..., K$ inequality constraints, $h_k(x) \leq 0$.

As before, we form the Lagrangean, putting a Lagrange multiplier λ on the equality constraint, and $\mu = (\mu_1, \mu_2, ..., \mu_K)$ *Kuhn-Tucker multipliers* on the inequality constraints:

$$\mathcal{L}(x, \lambda, \mu) = f(x) - \lambda g(x) - \sum_k \mu_k h_k(x) = f(x) - \lambda g(x) - \mu' h(x)$$

But now we have a dilemma, because some constraints may be binding at the solution while others are not. Moreover, there may be many solutions that correspond to different binding constraint sets. How are we supposed to know which bind and which don't? We proceed by making the set of FONCs larger to include *complementary slackness conditions.*

Theorem 11.1.1 (Kuhn-Tucker First-Order Necessary Conditions) *Suppose that*

1. *Suppose x^* is a local maximum of $f(x)$ subject to the constraints $g(x) = 0$ and $h_k(x) \leq 0$ for all k.*

2. *Let $E = \{1, 2, ..., \ell\}$ be the set of inequality constraints that* binding at *x^*, so that $h_j(x^*) = 0$ for all $j \in E$ and $h_j(x^*) < 0$ for all $j \notin E$. Suppose that the matrix formed by the gradient of the equality constraint, $\nabla g(x^*)$, and the binding inequality constraints in E, $\nabla h_1(x^*), \nabla h_2(x^*), ..., \nabla h_\ell(x^*)$, is non-singular (this is the constraint qualification).*

3. *The objective function, equality constraint, and inequality constraints are all differentiable at x^**

Then, there exist unique vectors λ^* and μ^* of multipliers so that

- *the FONCs*

$$\nabla f(x^*) - \lambda^* \nabla g(x^*) - \sum_j \mu_j^* h_j(x^*) = 0$$

$$-g(x^*) = 0$$

- *and the* complementary slackness *conditions*

$$\mu_j^* h_j(x^*) = 0$$

  for all $k = 1, ..., K$

are satisfied.

So, we always have the issue of the constraint qualification lurking in the background of a constrained maximization problem, so set that aside. The rest of the theorem says that if x^* is a local maximum subject to the equality constraint and inequality constraints, some will be active/binding and some will be inactive/slack. If we treat the active/binding constraints as regular equality constraints, then $\mu_j^* \geq 0$ for each binding constraint, but each slack constraint will satisfy $h_j(x^*) < 0$, so $\mu_j^* = 0$ — these are the complementary slackness conditions. (Note that it is not obvious right now that $\mu_j^* \geq 0$, but this is shown later.)

Example Suppose a consumer's utility function is $u(x, y)$, and he faces constraints $x \geq 0$, $y \geq 0$ and $w = px + y$. Suppose that $\nabla u(x, y) \geq 0$, so that the consumer's utility is weakly increasing in the amount of each good.

Then we can form the Lagrangian

$$\mathcal{L}(x, y, \mu) = u(x, y) - \mu_x x - \mu_y y - \lambda(px + y - w)$$

The FONCs are

$$u_x(x^*, y^*) - \mu_x^* - \lambda^* p = 0$$
$$u_y(x^*, y^*) - \mu_y^* - \lambda^* = 0$$
$$-(px^* + y^* - w) = 0$$

and the complementary slackness conditions are

$$\mu_x^* x^* = 0$$
$$\mu_y^* y^* = 0$$

Now we *hypothetically* have $2^2 = 4$ cases:

1. $\mu_x^*, \mu_y^* \geq 0$ (which will imply $x^* = 0, y^* = 0$ by the complementary slackness conditions)
2. $\mu_x^* = 0$,$\mu_y^* \geq 0$ (which will imply $x^* > 0, y^* = 0$ by the complementary slackness conditions)
3. $\mu_x^* \geq 0$,$\mu_y^* = 0$ (which will imply $x^* = 0, y^* > 0$ by the complementary slackness conditions)
4. $\mu_x^*, \mu_y^* = 0$ (which will imply $x^*, y^* > 0$ by the complementary slackness conditions)

We now solve the FONCs case-by-case:

1. In the first case, $\mu_x^*, \mu_y^* \geq 0$, we look at the corresponding complementary slackness conditions

$$\mu_x^* x^* = 0$$
$$\mu_y^* y^* = 0$$

   For these to hold, it must be the case that $x^* = 0$ and $y^* = 0$. Substituting these into the FONCs yields

$$u_x(0, 0) - \mu_x^* - \lambda^* p = 0$$
$$u_y(0, 0) - \mu_y^* - \lambda^* = 0$$
$$w = 0$$

   which yields a contradiction, since $w = \neq 0$. This case provides no candidate solutions.

2. In the second case, $\mu_x^* = 0$,$\mu_y^* \geq 0$, we look at the complementary slackness conditions

$$\mu_x^* x^* = 0$$

$$\mu_y^* y^* = 0$$

For these to hold, y^* must be zero. Substituting this into the FONCs yields

$$u_x(x^*, 0) - \lambda^* p = 0$$

$$u_y(x^*, 0) - \mu_y^* - \lambda^* = 0$$

$$-(px^* - w) = 0$$

from which we get $x^* = w/p$ from the final equation. This case gives a candidate solution $x^* = w/p$ and $y^* = 0$.

3. In the third case, $\mu_x^* \neq 0$,$\mu_y^* = 0$, we look at the complementary slackness conditions

$$\mu_x^* x^* = 0$$

$$\mu_y^* y^* = 0$$

For these to hold, we must have $x^* = 0$. Substituting this into the FONCs yields

$$u_x(0, y^*) - \mu_x^* - \lambda^* p = 0$$

$$u_y(0, y^*) - \lambda^* = 0$$

$$-(y^* - w) = 0$$

from which we get $y^* = w$ from the final equation. This case gives a candidate solution $x^* = 0$ and $y^* = w$.

4. In the last case, $\mu_x^*, \mu_y^* = 0$, the complementary slackness conditions are satisfied automatically. Substituting this into the FONCS yields

$$u_x(x^*, y^*) - \lambda^* p = 0$$

$$u_y(x^*, y^*) - \lambda^* = 0$$

$$-(px^* + y^* - w) = 0$$

which gives an interior solution $x^*, y^* > 0$.

So the KT conditions produce three candidate solutions: All x and no y, all y and no x, and some of each. Which solution is selected will depend on whether $\nabla u(x, y)$ is strictly greater than zero or only weakly greater than zero in each good. ∎

The complementary slackness conditions work as follows: Suppose we have an inequality constraint $x \leq 0$. Then the complementary slackness condition is that $\mu_x^* x^* = 0$, where μ_x^* is the Kuhn-Tucker multiplier associated with the constraint $x \leq 0$. Then if $\mu_x^* \geq 0$, it must be the case that $x^* = 0$ for the complementary slackness condition to hold; if $\mu_x^* = 0$, then x^* can take any value. This is how KT theory systematically works through all the possibilities of binding constraints: if the KT multiplier is non-zero, the constraint must be binding, while if the KT multiplier is zero, it is slack.

Example Consider maximizing $f(x,y) = x^{\alpha} y^{\beta}$ subject to the inequality constraints

$$\begin{aligned} y + 2x &\geq 6 \\ x + 2y &\geq 6 \\ x &\geq 0 \\ y &\geq 0 \end{aligned}$$

Then there are four inequality constraints. The set of points satisfying all four is not empty (graph it to check). However, the first two constraints only bind at the same time when $x = y = 2$; otherwise *either* the first is binding *or* the second but *not both*. This means that the Lagrangean is not the right tool, because you could only satisfy both equations as equality constraints at the point $x = y = 2$, which may not be a solution. You could set up a series of problems where you treat the constraints as binding or non-binding, and solve the 2^4 resulting maximization problems that come from considering each subset of constraints separately, but this is exactly what Kuhn-Tucker does for you, in a systematic way. The Lagrangean is

$$\mathcal{L} = \alpha \log(x) + \beta \log(y) - \mu_1(y + 2x - 6) - \mu_2(x + 2y - 6) - \mu_3 x - \mu_4 y$$

The first-order necessary conditions are

$$\begin{bmatrix} \partial \mathcal{L}/\partial x \\ \partial \mathcal{L}/\partial y \end{bmatrix} = \begin{bmatrix} \dfrac{\alpha}{x} - 2\mu_1 - \mu_2 - \mu_3 \\ \dfrac{\beta}{y} - \mu_1 - 2\mu_2 - \mu_4 \end{bmatrix}$$

The complementary slackness conditions are

$$\begin{aligned} \mu_1(y + 2x - 6) &= 0 \\ \mu_2(x + 2y - 6) &= 0 \\ \mu_3 x &= 0 \\ \mu_4 y &= 0 \end{aligned}$$

We now proceed by picking combinations of multipliers, and setting $\mu_j \geq 0$ if $h_j(x^*, y^*) = 0$, and $\mu_j = 0$ if $h_j(x^*, y^*) > 0$. This is what the "complementary" part of complementary slackness means.

First, note that the gradient is strictly increasing, so that if $x = 0$ or $y = 0$ at the optimium, it occurs at an extreme point of the set. So if $x = 0$ ($\mu_3 \geq 0$), then y must be as large as possible, and if $y = 0$ ($\mu_4 \geq 0$), x must be as large as possible. Second, note that if $6 = y + 2x$ is binding ($\mu_1 \geq 0$), then $6 = x + 2y$ is slack ($\mu_2 = 0$), and vice versa.

So we really have five possibilities:

1. $x = y = 2$ and both $\mu_1, \mu_2 \geq 0$ and $\mu_3, \mu_4 = 0$
2. $x, y > 0$ and $\mu_1 \geq 0$, $\mu_2, \mu_3, \mu_4 = 0$
3. $x, y > 0$ and $\mu_2 \geq 0$, $\mu_1, \mu_3, \mu_4 = 0$
4. $x > 0$ and $y = 0$, and $\mu_1 \geq 0$, $\mu_2, \mu_3, \mu_4 = 0$
5. $y > 0$ and $x = 0$, and $\mu_2 \geq 0$, $\mu_1, \mu_3, \mu_4 = 0$

For each case, we find the set of critical points that satisfy the resulting first-order conditions:

1. The unique critical point is $x = y = 2$.

2. If $x, y > 0$ and $\mu_1 \geq 0$, the binding constraint is that $y + 2x = 6$. The first-order necessary conditions become
$$\begin{bmatrix} \alpha/x - 2\mu_1 \\ \beta/y - \mu_1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$
This implies (why?) $\alpha y = 2\beta x$. With the constraint, this implies
$$x^* = \frac{3\alpha}{\alpha+\beta}, y^* = \frac{6\beta}{\alpha+\beta}$$

3. If $x, y > 0$ and $\mu_2 \geq 0$, the binding constraint is that $2y + x = 6$. The first-order necessary conditions become
$$\begin{bmatrix} \alpha/x - \mu_2 \\ \beta/y - 2\mu_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$
Using the conditions and constraint gives
$$x^* = \frac{6\alpha}{\alpha+\beta}, y^* = \frac{3\beta}{\alpha+\beta}$$

4. If $x > 0$ and $y^* = 0$, and $\mu_1 \geq 0$, then the constraint $2x + y = 6$ binds, and $x^* = 3$.

5. If $y > 0$ and $x^* = 0$, and $\mu_2 \geq 0$, then the constraint $x + 2y = 6$ binds, and $y^* = 3$.

So we have a candidate list of five potential solutions:

$$\begin{array}{rcl} 2 & , & 2 \\ \dfrac{6\alpha}{\alpha+\beta} & , & \dfrac{3\beta}{\alpha+\beta} \\ \dfrac{3\alpha}{\alpha+\beta} & , & \dfrac{6\beta}{\alpha+\beta} \\ 3 & , & 0 \\ 0 & , & 3 \end{array}$$

To figure out which is the global maximum, you generally need to plug these back into the objective and determine which point achieves the highest value. This will depend on the value of α and β. ∎

Example Suppose an agent is trying to solve an optimal portfolio problem, where two assets, x and y are available. Asset x has mean m_x and variance v_x, while asset y has mean m_y and variance v_y. The covariance of x and y is v_{xy}. He has wealth w, shares of x cost p, and shares of y cost 1, so his budget constraint is $w = xm_x + ym_y$, but $x \geq 0$ and $y \geq 0$, so he cannot short. The objective function is the mean minus the variance of the portfolio,
$$xm_x + ym_y - (x^2 v_x + ym_y^* - 2xyv_{xy})$$
Then the Lagrangean is
$$\mathcal{L} = xm_x + ym_y - (x^2 v_x + ym_y^* - 2v_{xy}) - \lambda(w - px - y) - \mu_x x - \mu_y y$$
The FONCs are
$$m_x - (2x^* v_x - 2y^* v_{xy}) - \lambda p - \mu_x = 0$$
$$m_y - (2y^* v_y - 2x^* v_{xy}) - \lambda - \mu_y = 0$$
$$-(w - px^* - y^*) = 0$$

and the complementary slackness conditions are

$$\mu_x^* x^* = 0$$

$$\mu_y^* y^* = 0$$

Then there are $2^2 = 4$ cases:

1. $\mu_x^* = mu_y^* = 0$, placing no restrictions on x^* and y^*, by the complementary slackness conditions
2. $\mu_x^* = 0$ and $\mu_y^* \geq 0$, so that $y^* = 0$, by the complementary slackness conditions
3. $\mu_x^* \geq 0$ and $\mu_y^* = 0$, so that $x^* = 0$, by the complementary slackness conditions
4. $\mu_x^* \geq 0$ and $\mu_y^* \geq 0$, so that $x^* = y^* = 0$, by the complementary slackness conditions

We solve case by case:

1. If $\mu_x^* = \mu_y^* = 0$, then the FONCs are

$$m_x - (2x^* v_x - 2y^* v_{xy}) - \lambda^* p = 0$$

$$m_y - (2y^* v_y - 2x^* v_{xy}) - \lambda^* = 0$$

$$-(w - px^* - y^*) = 0$$

   which is a quadratic programming problem and can be solved by matrix methods. In particular,

$$y^* = \frac{w(v_x + pv_{xy}) - \frac{p}{2}(m_x - pm_y)}{p(2v_{xy} + v_y) + v_x}$$

2. If $\mu_x^* = 0$ and $\mu_y^* \geq 0$, then the FONCs are

$$m_x - (2x^* v_x) - \lambda^* p = 0$$

$$m_y - (-2x^* v_{xy}) - \lambda^* - \mu_y^* = 0$$

$$-(w - px^*) = 0$$

   And the last equation gives a candidate solution $x^* = w/p$ and $y^* = 0$.

3. This case is similar to the previous one, with $y^* = w/p$ and $x^* = 0$.

So the portfolio optimization problem produces three candidate solutions. Our candidate solution with x^* and y^* both positive requires

$$w(v_x + pv_{xy}) - \frac{p}{2}(m_x - pm_y) > 0$$

If $p = 1$, this reduces to

$$w > \frac{m_x - m_y}{2(v_x + v_{xy})}$$

so that the difference in means can't be too large $(m_x - m_y)$ or the agent will prefer to purchase all x. If v_x or v_{xy} is large, however, this makes the agent prefer to diversify, even if $m_x > m_y$. The KT conditions allow us to pick out the various solutions and compare them, even for large numbers of assets (since this is just a quadratic programming problem). ∎

11.1.1 The sign of the KT multipliers

We haven't yet shown that $\mu_j \geq 0$. To do this, consider relaxing the first inequality constraint:

$$h_1(x) \leq \varepsilon$$

where ε is a relaxation parameter that makes the constraint easier to satisfy. The Lagrangian for a problem with one equality constraint and J inequality constraints becomes

$$\mathcal{L} = f(x) - \lambda g(x) - \mu_1(h_1(x) - \varepsilon) - \sum_{j=2}^{J} \mu_j h_j(x)$$

Let $V(\varepsilon)$ be the value function in terms of ε. If we have relaxed the constraint, the agent's choice set is larger, and he must be better off, so that $V'(\varepsilon) \geq 0$. But by the envelope theorem,

$$V'(\varepsilon) = \frac{\partial \mathcal{L}(x^*, \lambda^*, \mu^*)}{\partial \varepsilon} = \mu_1^* \geq 0$$

So the KT multipliers must be weakly positive, as long as your constraints are always written as $h_j(x) \leq 0$.

11.2 Second-Order Sufficient Conditions and the IFT

Some good news: Since KT maximization is essentially the same as Lagrange maximization once you have fixed the set of binding constraints, the SOSCs and implicit function theorem for an inequality-constrained problem are to those in an equality-constrained maximization problem where the set of binding constraints are equality constraints, and the slack constraints are completely ignored.

Remember: KT is just a formal way of working through the process of testing all the possible combinations of binding constraints. So for a particular set of binding constraints, the problem is equivalent to an *equality constrained* problem where these are the *only* constraints. To check whether a critical point is a local maximum or compute comparative statics, we use exactly the same approach as an equality constrained problem.

The envelope theorem is similar, but you have to be careful. In an inequality constrained problem there are potentially many local solutions for some values of the exogenous variables. But the envelope theorem is computed from the maximum of all of them, and as you vary the exogenous variables, you have to be careful that you don't drift into another solution's territory. Locally, there is not much to worry about, since you are almost never on such a boundary (in the sense of Lebesgue measure zero). But suppose you want to plot the maximized value of the objective function for an agent facing a savings behavior for a variety of interest rates where there is a borrowing constraint that binds for some values of the interest rate but not for others. However, this upper envelope over all maximization solutions is continuous, at least (but not necessary differentiable at "kinks" where the regime switches):

Theorem 11.2.1 (Theorem of the Maximum) *Consider maximizing $f(x, c)$ over x, where the choice set defined by the equality and inequality constraints is compact and potentially depends on c. Then $f(x^*(c), c)$ is a continuous function in c.*

The theorem of the maximum is a kind of generalization of the envelope theorem that says, "While $V(c)$ may not be differentiable because of the kinks where the set of binding constraints shifts, the value functions will always be at least continuous in a parameter c."

Exercises

1. Suppose an agent has the following utility function:

$$u(x_1, x_2) = \sqrt{x_1} + \sqrt{x_2}$$

and faces a linear constraint

$$1 = px_1 + x_2$$

and non-negativity constraints $x_1 \geq 0$, $x_2 \geq 0$.
i. Verify the objective function is concave in (x_1, x_2). ii. Graph the budget set for various prices p, and some of the upper contour sets. iii. Solve the agent's optimization problem for all $p > 0$. Check the second-order conditions are satisfied at the optimum. iv. Sketch the set of maximizers as a function of p, $x_1(p)$. v. Solve the agent's problem when $u(x_1, x_2) = \sqrt{x_1} + \sqrt{x_2 + 1}$. Is it possible to get corner solutions in this case? Explain the difference between the old objective function and the new one.

2. Suppose a firm hires capital K and labor L to produce output using technology $q = F(K, L)$, with $\nabla F(K, L) \geq 0$. The price of capital is r and the price of labor is w. Solve the *cost minimization problem*

$$\min_{K,L} rK + wL$$

subject to $F(K, L) \geq \bar{q}$, $K \geq 0$ and $L \geq 0$. Check the SOSCs or otherwise prove that your solutions are local maxima subject to the constraints. How does (K^*, L^*) vary in r and $\bar{q}$ at each solution? How does the firm's cost function vary in $\bar{q}$?

3. Warning: This question is better approached by thinking without a Lagrangean than trying to use KT theory. A manager has two factories at his disposal, a and b. The cost function for factory a is $C_a(q_a)$ where $C_a(0) = 0$ and $C_a(q_a)$ is increasing, differentiable, and convex. The cost function for factory b is

$$C_b(q_b) = \begin{cases} 0 & \text{if } q_b = 0 \\ c_b q_b + F & \text{if } q_b > 0 \end{cases}$$

i. Is factory b's cost function convex? Briefly explain the economic intuition of the property $C_b(0) = 0$ but $\lim_{q_b \downarrow 0} C_b(q_b) = F$. ii. Solve for the cost-minimizing production plan (q_a, q_b) that achieves $\bar{q}$ units of output, $q_a + q_b = \bar{q}$, $q_a \geq 0$, $q_b \geq 0$. Is the cost function $C^*(\bar{q})$ continuous? Convex? Differentiable? iii. Let p be the price of the good. If the firm's profit function is $\pi(q) = pq - C^*(q)$, what is the set of maximizers of $\pi(q)$?

4. Suppose an agent has the following utility function:

$$u(x_1, x_2) = \begin{cases} \sqrt{x_1 x_2} & , x_2 > x_1 \\ \dfrac{x_1 + x_2}{2} & , x_2 \leq x_1 \end{cases}$$

and faces a linear constraint

$$1 = px_1 + x_2$$

and non-negativity constraints $x_1 \geq 0$, $x_2 \geq 0$. (i.) Verify that these preferences are continuous but non-differentiable along the ray $x_1 = x_2$. (ii.) Graph the budget set for various prices p, and some of the upper contour sets. (iii.) Solve the agent's optimization problem for all $p > 0$. (iv.) Sketch the set of maximizers as a function/correspondence of p, $x_1(p)$.

5. Suppose an agent gets utility from consumption, $c \geq 0$, and leisure, $\ell \geq 0$. He has one unit of time, which can also be spent working, $h \geq 0$. From working, he gets a wage of w per hour, and his utility function is

$$u(c, \ell)$$

However, he faces taxes, which take the form

$$t(wh) = \begin{cases} 0 & , wh < t_0 \\ \tau wh & , wh \geq t_0 \end{cases}$$

Therefore, his income is wh, consumption costs pc, and he is taxed based on whether his wage is above or below a certain threshold, linearly at rate τ. i. Sketch the agent's budget set. Is it a convex set? ii. Formulate a maximization problem and characterize any first-order necessary conditions or complementary slackness conditions. iii. Characterize the agent's behavior as a function of w and t_0. What implications does this model have for the design of tax codes? iv. If the tax took the form

$$t(wh) = \begin{cases} 0 & , wh < t_0 \\ \tau & , wh \geq t_0 \end{cases}$$

sketch the agent's budget set. Is it convex? What implications does the geometry of this budget set have for consumer behavior?

6. A consumer has wealth w, and access to a riskless asset with return $R > 1$, and a risky asset with return $\tilde{r} \sim N(r, \sigma^2)$. He places ϕ_1 of his wealth into the riskless asset, and ϕ_2 of his wealth into the risky asset:

$$w = \phi_1 + \phi_2$$

and he maximizes the mean-less-the-variance of his returns:

$$U(\phi_1, \phi_2) = \phi_1 R + \phi_2 r - \frac{\gamma}{2}\phi_2^2(\sigma^2 + r^2)$$

(i) Find conditions on $\lambda, R, r^*, \sigma^2$ so the agent holds some of both the risky and riskless assets, and solve for the optimal portfolio weights, (ϕ_1^*, ϕ_2^*). (ii) How do the optimal portfolio weights vary with r^*, σ^2 and R? (iii) How does his payoff vary with w and R when he uses the optimal portfolio?

Chapter 12

Concavity and Quasi-Concavity

Checking second-order sufficient conditions for equality- and inequality-constrained maximization problems is often outrageously painful. It is time-consuming and error-prone, and pain increases exponentially in the number of choice variables.

For this reason, it would be very helpful to know when the second-order sufficient conditions for multi-dimensional maximization problems are automatically satisfied.

There are two classes of function that are useful for maximization: Concave and Quasi-concave. (The corresponding classes for minimization are Convex and Quasi-convex.)

In the one-dimensional case, strict concavity meant that $f''(x) < 0$ for all x. In the multi-dimensional case, strictly concavity will similarly mean that $y'\nabla_{xx}f(x)y < 0$ for all x and y, and it follows that the second-order conditions will be satisfied. This is the best case.

However, recall that if x^* is a local maximum of $f(x)$, then for any strictly increasing function $g()$, x^* is also a maximum of $g(f(x))$. This is nice, because it means that the "units" of $f(x)$ are irrelevant to the set of maximizers. But concavity and convexity are *not* preserved under monotone transformations. For example, $\log(x)$ is one of our prototype concave functions. But if we take the strictly increasing transformation $g(y) = (e^y)^2$, we get

$$g(\log(x)) = x^2$$

which is a convex function.

This has the potential to cause problems for us as economists, because we want the set of maximizers to be independent of how we describe $f(x)$ up to increasing transformations (that lets us claim that we don't need "utils" or *cardinal utility* to measure people's preferences, we can just observe how they behave). We want our theories to be based on *ordinal utility*, meaning that the numbers provided by a utility function are meaningless in themselves, and only serve to verify whether one option is better than another.

This is where quasi-concavity comes in. It is a property similar to concavity that is preserved under monotone transformations, and has the same convenient theorems built in.

12.1 Some Geometry of Convex Sets

For analyzing maximization problems, it is helpful to consider the behavior of the "better than" sets:

Definition 12.1.1 *The* upper contour sets *of a function are the sets*

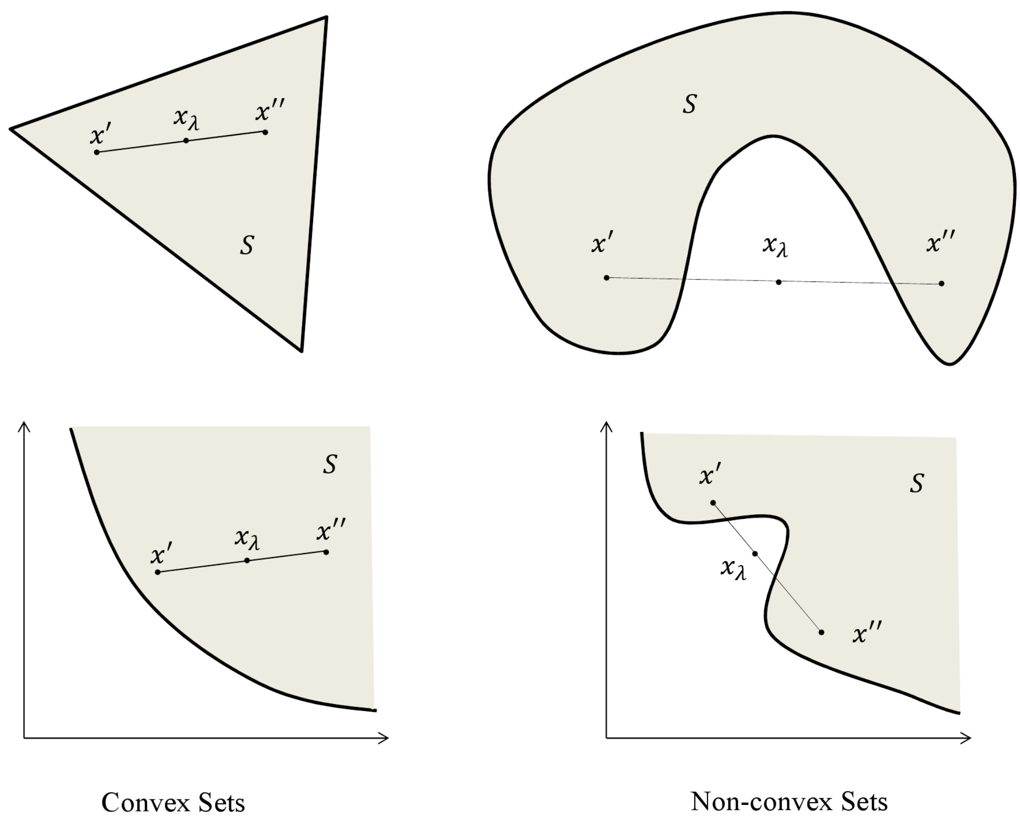
$$UC(a) = \{x : f(x) \geq a\}$$

Now, we need to be careful about the difference between convex *sets* and convex *functions*, since we're about to use both words side by side quite a bit.

Definition 12.1.2 *A set S is* convex *if, for any x' and x'' in S and any $\lambda \in [0,1]$, the point*

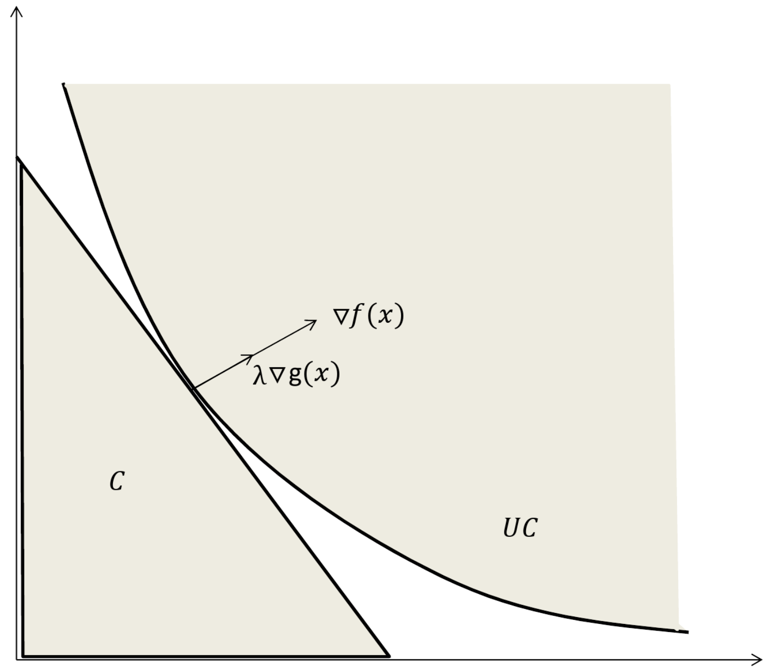
$$x_\lambda = \lambda x' + (1-\lambda)x''$$

is also in S. The interior *of a convex set S, $\int(S)$, are all the points x for which there is a ball $B_\delta(x) \subset S$. If any convex combination is in the interior of S, then S is a* strictly convex *set.*



Convexity

Convex sets are well behaved since you can never drift outside the set along a chord between two points. Geometrically, if your upper contour sets are convex and the constraint set is strictly convex, then we should be able to “separate” them, like this:



Separation of Constraint and Upper-contour Sets

This would mean there is a unique maximizer, since there is a unique point of tangency between the indifference curves and the constraint set.

Why? If we pick a point x_0 on an indifference curve and some point x_1 better than x_0, all of the options along the path from x_0 to x_1 are better than x_0. We might express this more formally by saying that for $\lambda \in [0,1]$, the options $x_\lambda = \lambda x_1 + (1-\lambda)x_0$ are all better than x_0, or that the upper contour sets are convex sets. This is a key geometric feature of well-behaved constrained maximization problems.

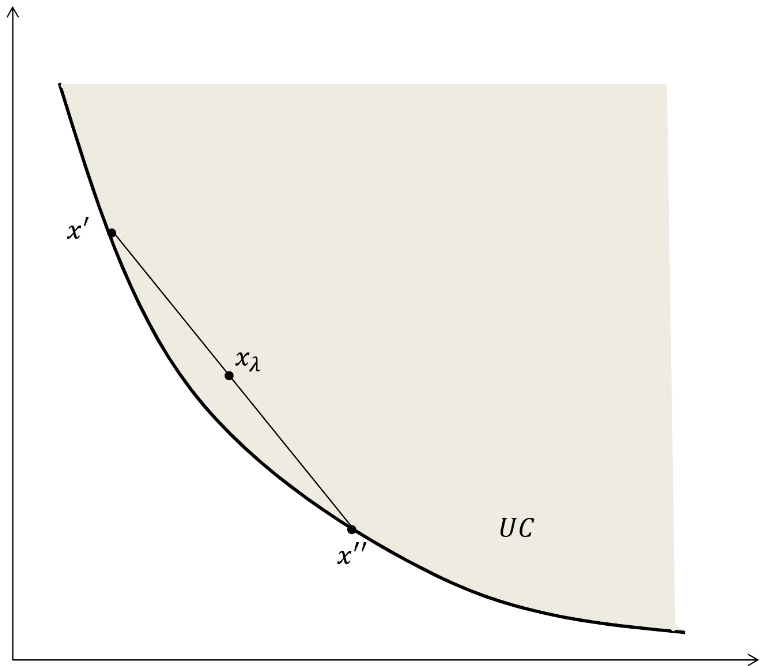
12.2 Concave Programming

Concave functions are the best-case scenario for optimization, since they imply that first-order necessary conditions are sufficient for a critical point to be a *global* maximum (this makes checking second-order sufficient conditions unnecessary, and life is much easier).

Definition 12.2.1 *A function $f(x)$ is* concave *if for all x' and x'' and all $\lambda \in [0,1]$,*

$$f(\lambda x' + (1-\lambda)x'') \geq \lambda f(x') + (1-\lambda)f(x'')$$

It is strictly concave *if the equality is strict for all $\lambda \in (0,1)$.*



Concave functions

This is a natural property in economics: Agents often prefer variety — consuming the bundle $\lambda x' + (1-\lambda)x''$ — to consuming either of two "extreme" bundles. For example, which sounds better: Five apples and five oranges, or ten apples with probability 1/2 and ten oranges with probability 1/2? If you say, "Five apples and five oranges", you have concave preferences.

However, there are many, equivalent ways to characterize a concave function:

Definition 12.2.2 *The following are equivalent:*

- *$f(x)$ is concave*
- *For all x' and x'' and all $\lambda \in (0,1)$,*
$$f(\lambda x' + (1-\lambda)x'') \geq \lambda f(x') + (1-\lambda)f(x'')$$
- *For all x' and x'',*
$$f(x') - f(x'') \leq \nabla f(x'')(x' - x'')$$
- *The Hessian of $f(x)$ satisfies $y'H(x)y \leq 0$ for all x in the domain of $f(x)$ and $y \in \mathbb{R}^N$.*

If the weak inequalities above are replaced with strict inequalities, the function is strictly concave.

The Hessian characterization — concave if $H(x)$ is negative semi-definite for all x in the domain of $f(x)$, and strictly concave if $H(x)$ is negative definite for all x in the domain of $f(x)$ — is extremely useful, since the Hessian is so ubiquitous in proving SOSCs. If the objective function is strictly concave, then at a critical point x^*,

$$f(x) = f(x^*) + (x - x^*)' \frac{H(x^*)}{2}(x - x^*) + o(h^3)$$

or for x close to x^*,

$$f(x^*) - f(x) = -(x - x^*) \frac{H(x^*)}{2}(x - x^*) > 0$$

since $y'H(x)y < 0$ for *any* x, including x^*. Then we can conclude that $f(x^*) > f(x)$, so that a critical point x^* is a local maximum.

Theorem 12.2.3 • *Consider an unconstrained maximization problem* $\max_x f(x)$. *If* $f(x)$ *is concave, any critical point is a global maximum of* $f(x)$. *If* $f(x)$ *is strictly concave, any critical point is the unique global maximum.*

- *Consider an inequality constrained maximization problem* $\max_x f(x)$ *subject to* $g(x) = 0$ *and for* $k = 1, ..., K$, $h_k(x) \leq 0$. *If* $f(x)$ *is concave and the constraint set is convex, any critical point of the Lagrangian is a global maximum. If* $f(x)$ *is strictly concave and the constraint set is convex, any critical point of the Lagrangian is a unique global maximum.*

Note that for Kuhn-Tucker, the situation is slightly more complicated than it might appear, since the objective function might be strictly concave so that for each set of active constraints, there is a unique global maximizer (since each is just an equality-constrained maximization problem), but these candidates must still be somehow compared.

12.3 Quasi-Concave Programming

Concavity is a cardinal property, not an ordinal one, and we would like to have a generalization of concavity that is merely ordinal.

For example, if we consider $x^\alpha y^\beta$ and compute the Hessian, we get

$$\begin{bmatrix} \alpha(\alpha-1)x^{\alpha-2}y^\beta & \alpha\beta x^{\alpha-1}y^{\beta-1} \\ \alpha\beta x^{\alpha-1}y^{\beta-1} & \beta(\beta-1)x^\alpha y^{\beta-2} \end{bmatrix}$$

which is not concave if $\alpha + \beta > 1$. But $f(x, y) = xy$ still has convex upper contour sets, so we would expect it to be well-behaved for maximization purposes, even if it isn't concave.

To deal with this, we introduce a new kind of function, quasi-concave functions. We're going to motivate quasi-concavity in a somewhat roundabout way. Recall that for a function $f(x, y)$, the indifference curve $x(y)$ was defined as the implicit solution to the equation

$$f(x(y), y) = c$$

Now, if the indifference curves $x(y)$ were

1. Downward sloping, so that taking some y away from the agent required increasing x to keep them on the same indifference curve

2. A convex function, so that taking away a lot of y away from the agent requires giving them ever higher quantities of x to compensate him

3. Invariant to strictly increasing transformations

we would have the right geometric properties for maximization without the cardinal baggage that comes with assuming concavity.

When are 1-3 satisfied? Well, for a strictly increasing transformation $g()$, $g(f(x(y), y))$ has indifference curves defined by

$$g'(f(x(y), y))(f_x(x(y), y)x'(y) + f_y(x(y), y)) = 0$$

which are the same as those generated by $f(x(y), y)$. So far so good. Now, the derivative of $x(y)$ is

$$x'(y) = -\frac{f_y(x(y), y)}{f_x(x(y), y)}$$

And the second derivative is

$$x''(y) = -\frac{(f_{xy}x'(y) + f_{yy})f_x - (f_{xx}x'(y) + f_{yx})f_y}{f_x^2} > 0$$

which implies

$$\left(-f_{xy}\frac{f_y}{f_x} + f_{yy}\right) f_x - \left(-f_{xx}\frac{f_y}{f_x} + f_{yx}\right) f_y < 0$$

or

$$-f_{xy}f_yf_x + f_{yy}f_y^2 + f_x^2 f_{xx} - f_{xy}f_yf_x < 0$$

Which doesn't appear to be anything at first glance. Actually, it is the determinant of the matrix

$$\begin{bmatrix} 0 & f_x & f_y \\ f_x & f_{xx} & f_{xy} \\ f_y & f_{yx} & f_{yy} \end{bmatrix}$$

So if this matrix is negative semi-definite, the indifference curves $x(y)$ will be concave functions, and if it is negative definite, the indifference curves $x(y)$ will be strictly concave functions. This is the idea of quasi-concavity.

Definition 12.3.1 *The following are equivalent:*

- $f(x)$ *is quasi-concave*
- *For every real number* a*, the upper contour sets of* $f(x)$*,*

  $$UC(a) = \{x : f(x) \geq a\}$$

  are convex.
- *The bordered Hessian*

  $$H(x) \begin{bmatrix} 0 & \nabla_x f(x)' \\ \nabla_x f(x) & \nabla_{xx} f(x) \end{bmatrix}$$

  is negative semi-definite, $y'H(x)y \leq 0$ *for all* y*.*
- *For all* λ *in* $[0, 1]$*,*

  $$f(\lambda x' + (1-\lambda)x'') \geq \min\{f(x'), f(x'')\}$$
- *If* $f(x') \geq f(x'')$*, then*

  $$\nabla f(x'')(x' - x'') \geq 0$$

If the weak inequalities above are replaced with strict inequalities, the function is strictly quasi-concave*.*

Quasi-concavity is useful because of the following theorem:

Theorem 12.3.2 *Consider an inequality constrained maximization problem* $\max_x f(x)$ *subject to* $g(x) = 0$ *and for* $k = 1, ..., K$*,* $h_k(x) \leq 0$*. If* $f(x)$ *is quasi-concave and the constraint set is convex, any critical point of the Lagrangian is a global maximum. If* $f(x)$ *is strictly concave and the constraint set is convex, any critical point of the Lagrangian is a unique global maximum.*

Note that in unconstrained problems, quasi-concavity might not be enough to guarantee a critical point is a maximum. For example, the function $f(x) = x^3$ is quasi-concave, since the upper contour sets $UC(a) = \{x : x^3 \geq a\}$ are convex sets. However, it fails to achieve a maximum.

To check that a function is quasi-concave, let

$$H_k(x) = \begin{bmatrix} 0 & \partial f(x)/\partial x_1 & \partial f(x)/\partial x_2 & \dots & \partial f(x)/\partial x_{k-1} \\ \partial f(x)/\partial x_1 & \partial^2 f(x)/\partial x_1^2 & \partial^2 f(x)/\partial x_2 \partial x_1 & \dots & \partial^2 f(x)/\partial x_1 \partial x_{k-1} \\ \partial f(x)/\partial x_2 & \partial^2 f(x)/\partial x_1 \partial x_2 & \partial^2 f(x)/\partial x_2^2 & \dots & \partial^2 f(x)/\partial x_2 \partial x_{k-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \partial f(x)/\partial x_{k-1} & \partial^2 f(x)/\partial x_{k-1} \partial x_1 & \partial^2 f(x)/\partial x_{k-1} \partial x_2 & \dots & \partial^2 f(x)/\partial x_{k-1}^2 \end{bmatrix}$$

Then we have an alternating sign test on the determinants of the principal minors:

Theorem 12.3.3 (Alternating Sign Test) *If f is quasiconcave, then the determinants of the leading principal minors of the bordered Hessian alternate in sign, starting with $\det H_3(x) \geq 0$, $\det H_4(x) \leq 0$, and so on. If the leading principal minors of the bordered Hessian alternate in sign, starting with $\det H_3(x) > 0$, $\det H_4(x) < 0$, and so on, then f is quasiconcave.*

12.4 Convexity and Quasi-Convexity

The set of convex and quasi-convex functions play a similar role in minimization theory as concave and quasi-concave functions play in maximization theory.

Definition 12.4.1 *The following are equivalent:*

- *$f(x)$ is convex*
- *For all x' and x'' and all $\lambda \in [0, 1]$,*

$$\lambda f(x') + (1 - \lambda) f(x'') \leq f(\lambda x' + (1 - \lambda) x'')$$

- *For all x' and x'',*

$$f(x') - f(x'') \geq Df(x'')(x' - x'')$$

- *The Hessian of $f(x)$ satisfies $y'H(x)y \geq 0$ for all x in the domain of $f(x)$ and $y \in \mathbb{R}^N$.*

If the weak inequalities above are replaced with strict inequalities, the function is strictly convex.

and

Definition 12.4.2 *The following are equivalent:*

- *$f(x)$ is quasi-convex*
- *For every real number a, the lower contour sets of $f(x)$,*

$$LC(a) = \{x : f(x) \leq a\}$$

  are convex.

- *The bordered Hessian*

$$H(x) \begin{bmatrix} 0 & \nabla_x f(x)' \\ \nabla_x f(x) & \nabla_{xx} f(x) \end{bmatrix}$$

  is positive semi-definite, $y'H(x)y \geq 0$ for all y.

- *For all* λ *in* $(0,1)$,

$$f(\lambda x' + (1-\lambda)x'') \leq \max f(x'), f(x'')\}$$

- *If* $f(x') \leq f(x'')$, *then*

$$\nabla f(x'')(x' - x'') \leq 0$$

If the weak inequalities above are replaced with strict inequalities, the function is strictly quasi-convex.

and we have

Theorem 12.4.3 • *Consider an unconstrained maximization problem* $\max_x f(x)$. *If* $f(x)$ *is convex, then any critical point of* $f(x)$ *is a global minimizer of* $f(x)$. *If* $f(x)$ *is strictly convex, any critical point is the unique global minimum.*

- *Consider an equality constrained maximization problem* $\max_x f(x)$ *subject to* $g(x) = 0$, *where* $f(x)$ *and* $g(x)$. *If* $f(x)$ *is convex or quasi-convex, any critical point of the Lagrangian is a global maximum. If* $f(x)$ *is strictly convex or strictly quasi-convex and the constraint set is convex, any critical point of the Lagrangian is a unique global minimum.*

- *Consider an inequality constrained maximization problem* $\max_x f(x)$ *subject to* $g(x) = 0$ *and for* $k = 1, ..., K$, $h_k(x) \leq 0$. *If* $f(x)$ *is convex or quasi-convex, any critical point of the Lagrangian is a global minimum. If* $f(x)$ *is strictly convex or strictly quasi-convex and the constraint set is convex, any critical point of the Lagrangian is a unique global minimum.*

The alternating sign test for quasi-convexity is slightly different:

Theorem 12.4.4 (Alternating Sign Test) *If* f *is quasiconvex, then the determinants of the leading principal minors of the bordered Hessian are all weakly negative. If the leading principal minors of the bordered Hessian are all strictly negative, then* f *is quasiconvex.*

12.5 Comparative Statics with Concavity and Convexity

There are some common tricks to deriving comparative statics relationships under the assumptions of concavity and convexity.

Example Suppose a firm chooses inputs $z = (z_1, z_2)$ whose costs per unit are $w = (w_1, w_2)$, subject to a production constraint $F(z_1, z_2) = q$. Let

$$c(w, q) = \min_z w \cdot z$$

subject to $F(z) = q$.

First, note that $c(w, q)$ is concave in w. How do we prove this? Suppose z' minimizes costs at w' and z'' minimizes costs at w''. Now consider the price vector $w_\lambda = \lambda w' + (1-\lambda)w''$, and its cost-minimizing solution z_λ. Then

$$c(w_\lambda, q) = w_\lambda z_\lambda = (\lambda w' + (1-\lambda)w'') \cdot z_\lambda = \lambda w' \cdot z_\lambda + (1-\lambda)w'' \cdot z_\lambda$$

But by definition $w' \cdot z_\lambda \geq w' \cdot z'$ and $w'' \cdot z_\lambda \geq w'' \cdot z''$, so that

$$w_\lambda z_\lambda \geq \lambda w' \cdot z' + (1-\lambda)w'' \cdot z''$$

or

$$c(\lambda w' + (1-\lambda)w'', q) \geq \lambda c(w', q) + (1-\lambda)c(w'', q)$$

so that the cost function is concave.

Second, we use the envelope theorem to differentiate the cost function with respect to w:

$$\nabla_w c(w,q) = z(w,q)$$

or

$$\begin{bmatrix} c_{w_1}(w_1,w_2,q) \\ c_{w_2}(w_1,w_2,q) \end{bmatrix} = \begin{bmatrix} z_1(w_1,w_2,q) \\ z_2(w_1,w_2,q) \end{bmatrix}$$

and again with respect to w to get

$$\nabla_{ww} c(w,q) = \nabla_w z(w,q)$$

or

$$\begin{bmatrix} c_{w_1w_1}(w_1,w_2,q) & c_{w_2w_1}(w_1,w_2,q) \\ c_{w_1w_2}(w_1,w_2,q) & c_{w_2w_2}(w_1,w_2,q) \end{bmatrix} = \begin{bmatrix} \dfrac{\partial z_1(w_1,w_2,q)}{\partial w_1} & \dfrac{\partial z_1(w_1,w_2,q)}{\partial w_2} \\ \dfrac{\partial z_2(w_1,w_2,q)}{\partial w_1} & \dfrac{\partial z_2(w_1,w_2,q)}{\partial w_2} \end{bmatrix}$$

Finally, since $c(w,q)$ is concave, every element on its diagonal is weakly negative. Therefore, every element on the diagonal of $z(w,q)$ must also be negative, or

$$\frac{\partial z_k(w,q)}{\partial w_k} = \frac{\partial^2 c(w_1,w_2,q)}{\partial w_k^2} \leq 0$$

so that if $w_k \uparrow$ then $z_k \downarrow$. ■

This exercise would actually be pretty difficult without the knowledge that $c(w,q)$ is concave, and that is the key.

Example Suppose a consumer buys bundles $q = (q_1, ..., q_N)$ of goods, has utility function $u(q,m) = v(q) + m$, and budget constraint $w = p'q + m$. Suppose $v(q)$ is concave. Then the objective function is

$$\max_q v(q) - p'q + w$$

The FONCs are

$$\nabla_q v(q^*) - p = 0$$

and totally differentiating with respect to p yields

$$\nabla_{qq} v(q^*) \nabla_p q^* - 1 = 0$$

and

$$\nabla_p q^* = [\nabla_{qq} v(q^*)]^{-1}$$

Since $v(q)$ is concave, its Hessian is negative definite, and the inverse of a negative definite matrix is negative definite, so all the entries on the diagonal are weakly negative. Therefore,

$$\frac{\partial q_k^*(p)}{\partial p_k} \leq 0$$

■

Again, the concavity of the objective function is what makes the last part do-able. Without knowing that $[\nabla_{qq} v(q^*)]^{-1}$ is negative semi-definite by the concavity of $v(q)$, we would be unable to decide the sign of the comparative statics of each good with respect to a change in its own-price.

Exercises

1. Let $(x, y) \in \mathbb{R}^2_+$. (i) When is $f(x, y) = x^\alpha y^\beta$ concave? Quasi-concave? (ii) When is $f(x, y) = \left(x^\rho + y^\rho\right)^{1/\rho}$ concave? (iii) When is $\min\{ax, by\}$ concave? Quasi-concave?

2. Show that if $f(x)$ is quasi-concave and $h(y)$ is strictly increasing, then $h(f(x))$ is quasi-concave. Show that x^* maximizes $f(x)$ subject to $h(x) = 0$ iff x^* maximizes $h(f(x))$ subject to $g(x) = 0$.

3. Show that any increasing function $f : \mathbb{R} \to \mathbb{R}$ is quasi-concave, but not every increasing function is concave. Show that the sum of two concave functions is concave, but the sum of two quasi-concave functions is not necessarily quasi-concave.

4. Prove that if $f(x)$ is convex, then any critical point of $f(x)$ is a global minimum of the unconstrained maximization problem.

5. Show that all concave functions are quasi-concave. Show that all convex functions are quasi-convex.

6. Suppose a firm maximizes profits $\pi(q, K, L) = pq - rK - wL$ subject to the constraint where $F(K, L) = q$. Show that $\pi()$ is convex in (p, r, w). Prove that the gradient of π with respect to (p, r, w) is equal to $(q, -K, -L)$. Lastly, show that q^* is increasing in p, K^* is decreasing in r, and L^* is decreasing in w. Note how these comparative statics do not rely on the properties of $F(K, L)$.